Identification of Panel Data Models with Endogenous Censoring*

S. Khan
Department of Economics
Duke University
Durham, NC - USA
shakeebk@duke.edu

M. Ponomareva
Department of Economics
Northern Illinois University
DeKalb, IL - USA
mponomareva@niu.edu

E. Tamer
Department of Economics
Northwestern University
Evanston, IL - USA
tamer@northwestern.edu

First version: December 2010. This version: December 17, 2013

Abstract

We study inference on parameters in censored panel data models, where the censoring can depend on both observable and unobservable variables in arbitrary ways. Under some general conditions, we characterize the information the model and data contain about the parameters of interest by deriving the identified sets: Every parameter that belongs to these sets is *observationally equivalent* to the true parameter - the one that generated the data . We consider two separate sets of assumptions (2 models): the first uses stationarity on the unobserved disturbance terms. The second is a *nonstationary* model with a conditional independence restriction. Based on the characterizations of the identified sets, we provide a valid inference procedure that is shown to yield correct confidence sets based on inverting stochastic dominance tests. Also, we also show how our results extend to empirically interesting dynamic versions of the model with both lagged observed outcomes, and lagged indicators. We also show extensions to models with factor loads. In addition, and for both models, we provide sufficient conditions for point identification in terms of support conditions.The paper then examines sizes of the identified sets, and a Monte Carlo exercise shows reasonable small sample performance of our procedures.

Keywords: Panel Data, Dependent Censoring, Partial Identification.

*We thank Sofya Budoneva and Marcelo Ochoa and Yichong Zhang for research assistance on this project, seminar participants at various institutions for comments, and the National Science Foundation for research support.

1 Introduction

We consider the problem of inference on β in the panel data model

$$y^*_{it} = \alpha_i + x'_{it}\beta + \epsilon_{it}, \quad t = 1, \ldots T \quad i = 1, \ldots, N$$

where α_i is an individual specific and time-independent fixed (or random) effect that is allowed to be correlated with both $\mathbf{x}_i = (x_{i1}, \ldots, x_{iT})$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})$. The outcome variable, y^*_{it}, is only observed when it is greater than a censoring variable c_{it} where c_{it} itself is only observed when it exceeds y^*_{it}. The censoring variable $\mathbf{c}_i = (c_{i1}, \ldots, c_{iT})$ is allowed to depend on $\boldsymbol{\epsilon}_i$ in an arbitrary way. We summarize this as follows:

$$\begin{aligned} \text{we observe for } i\text{:} \quad & (y_{it} = \max(y^*_{it}, c_{it}),\ 1[y^*_{it} \geq c_{it}],\ x_{it}) \quad t = 1, \ldots, T \\ \text{where} \quad & \boldsymbol{\epsilon}_i \not\perp\!\!\!\perp \mathbf{c}_i | \mathbf{x}_i \end{aligned}$$

The presence of this *endogenous* censoring represents a challenge for existing methods[1] that are used for correcting for censoring since these methods usually assume that $\mathbf{c}_i$ is either observed or (conditionally) independent of the errors. There, the observed censoring is motivated via some design or data limitation issue (such as top-coding), and hence is assumed independent of the outcome. Here, the starting point is we want to allow for this censored variable c_{it} to be on equal footing as the outcome and so allow it to be arbitrarily correlated with y^*_{it} (but also accommodate fixed and independent censoring[2]). This enlarges the set of models that are covered to include Roy-like competing risks models, switching regression like models, and duration models with attrition that are important in applied work[3].

In addition, we consider identification also for a set of *dynamic models* where we allow past outcomes to impact current outcomes. For example, we study models with some forms of dynamics:

$$y^*_{it} = \alpha_i + x'_{it}\beta + G(y^*_{i,t-1}; c_{it}) + \epsilon_{it}$$

[1]This is especially the case when T is finite -typically $T = 2$-, as we assume throughout this paper.

[2]In the cross sectional setting this model is popular in duration analysis, as it relates to the Accelerated Failure Time (AFT) model. See, e.g Khan and Tamer (2009) for more on this for cross sectional data. In the panel data setting considered in this paper, t does not refer to the time period, but the spell in question.

[3]A canonical empirical example of this kind of censoring is a wage panel regression with an indicator dummy of whether individual i belongs to a union in time t. The censoring occurs since a union member's nonunion wage in period t is censored but is presumed less than the observed union wage.

where we let 1) $G(y^*_{it-1}; c_{it}) = \gamma y_{i,t-1}$ -a model with lagged *observed* outcomes, 2) $G(y^*_{it-1}; c_{it}) = \gamma d_{i,t-1}$ - a model with lagged *censoring indicator*, and 3) $G(y^*_{it-1}; c_{it}) = \gamma y^*_{i,t-1}$ -a model with lagged latent outcome- and analyze the inference question in these models. In addition, we also consider the identification in a model with *time varying factor loads* whereby the fixed effects enters as α_i in the first period, whereas in period 2, the fixed effect enters as $\gamma\alpha_i$ (and so on if $T > 2$). This allows returns to unobservable "skills" to change over time. See Section 4 for details on these models.

Generally, point identification conditions in nonlinear panel data[4] models with fixed effects are often strong, partly since simple differencing techniques, used in linear models, are not available when the model is nonlinear in the unobserved individual specific variable. So, typical point identification strategies have relied on distributional assumptions, and/or support conditions that are problem specific that often times rule out economically relevant models and behaviors. This has motivated a complementary approach to inference in these models that recognizes the fact that though point identification might not be possible under weaker assumptions, these models do contain nontrivial information about β. So, instead of looking for conditions under which point identification is guaranteed, we posit a model for the data generating process and then analyze the question of what information does this model have about β given the observed data.

The challenge in this *bounds approach* to identification analysis is to consider all the information in the data and the model: that is, find the *tightest* bound that contains the set of observationally equivalent parameter values. So, we analyze the question of what can one learn about β under 2 sets of weak assumptions that generally do not point identify β. The main results in the paper show how one can construct *sharp sets* for β: there is no more information that the data contain about distinct β's in the identified set given the model assumptions, i.e., every parameter vector in these sets is observationally equivalent to the true parameter β under the model assumptions. This analysis allows us to determine under what conditions for example this set is the *trivial set* (data contain no information about β) on the one hand, or also examine when this set shrinks to a singleton, β. The usefulness in this approach is that we posit the model (or sets of assumptions) first and then ask what information do these assumptions contain about β

[4]For the latest on methods for linear panel models in econometrics, see Arellano (2003).

as opposed to the complementary approach based on point identification in which one looks for a model (a set of assumptions) that guarantees point identification under the weakest set of assumptions. Given this characterization of the identified set, we provide valid inference approaches that allow us to build confidence regions for the true β.

There are a set of recent papers that deal with various nonlinearities in models with (short T) panels. See for example the work of Arellano and Bonhomme (2009), Bester and Hansen (2009), Bonhomme (2012), Chernozhukov, Fernandez-Val, Hahn, and Newey (2010), Evdokimov (2010), Graham and Powell (2012) and Hoderlein and White (2012). An important early work is the paper by Honoré (1992) which considers a panel model with fixed censoring. See also the survey in Arellano and Honoré (2001).

Censored models play an important role in applied economics with panel data. The models of the kind we consider here can be seen as a panel extension of the classic Roy model (or switching regression model) where in every period, one chooses to work in one of two sectors and this decision is based on whether the wage in the one sector is higher than the wage in the other sector. It is crucial here to allow for endogenous censoring since (unobserved) determinants of wage in one sector will effect the potential wage in the other sector. Our model of censoring is also an example of a competing risks model that is well studied in both economics (see for example the recent work of Honoré and Lleras-Muney (2006)) and statistics where it is widely applied. Also, this model can be used in duration analysis where outcomes such as unemployment spells can be censored. Our setup also handles models with attrition in which outcomes are missing nonrandomly[5]. Finally, censoring can also be a result of mechanical considerations such as top-coding, and there, typically, the censoring is fixed (and hence exogenous - see Honoré (1992)). In addition, our methods can be used in models that include dynamics, such as lagged outcome variables or lagged sector specific variables as regressors, and also models with time varying factor loads.

Generally, missing or interval outcome models were considered in a nonparametric setup in the partial identification literature with cross section data. Manski and Tamer (2002) considered inference on the slope vector in a linear model with interval outcomes

[5] A typical situation is one where outcomes that are supposed to be measured at various prespecified times, but then, individuals are lost, or do not show up. This would be a model of attrition in which we observe the outcome until the individual drops out. Here, c_t is equal to $-\infty$ if the individual shows up, and switches to $+\infty$ when there is attrition.

using a partial identification approach. With panel data, Honoré and Tamer (2006) considered bounds on parameters of interest in some interesting nonlinear panel models. In this paper, our starting point is the panel model with endogenous censoring under two sets of maintained assumptions (we consider both stationary and non-stationary time and individual-specific errors). Our goal is to take assumptions that have been previously used in the literature to obtain point identification (fixed censoring), but now allow for arbitrary censoring that can be correlated with both the outcomes and the covariates, with arbitrary individual unobserved time-invariant heterogeneity (fixed effects). On the other hand, weakening the assumptions even further can result in the identification becoming trivial: any possible vector of parameters is consistent with the distribution of observables. A similar trade-off is shown, for example, by Rosen (2012) for quantile panel data models with fixed effects and small T. In particular, under a conditional median independence assumption on ϵ_{it}, Rosen (2012) showed that a linear panel model (with no censoring) contains no information on the true parameter β, so that the identified set is the whole parameter space. This happens because ϵ_{i1} is allowed to be arbitrarily correlated with ϵ_{i2} under the conditional median independence assumption.

The first set of assumptions we employ (**Model 1**) uses *stationarity* on the distribution of ϵ_{it}, but otherwise leaves the error distribution unconstrained (and hence allow for cross sectional heteroskedasticity). Stationarity in nonlinear panel models has been used extensively before since the work of Manski (1987) where there it was shown that the binary choice panel model point identifies β under a stationarity and a set of support conditions.

The second set of assumptions (**Model 2**) relaxes stationarity but instead imposes independence between $\boldsymbol{\epsilon}_i$ and $\mathbf{x}_i$. This non-stationary setup allows for the distribution of the error terms to vary arbitrarily across time periods. Again, we construct another set of conditional moment inequalities that is shown to sharply characterize the identified set under the above non-stationarity assumption. Using the structure of those inequalities, one can obtain conditions under which the model contains no information on the parameter of interest.

Finally, for both Model 1 and Model 2, we provide sufficient conditions for the identified set to be equal to $\{\beta\}$ (i.e. point identification). In addition, we show how our methods can be extended to allow for some kinds of dynamics in the model by accommodating lagged censored and latent dependent variable, and lagged indicators of censoring. In what follow, we assume that $T = 2$. The results we obtain can readily be extended to a case of any fixed $T \geq 2$.

Although the focus of the paper is the study of identification and characterization of information on β under generalized censoring, the conditional moment inequality restrictions that we construct to characterize this information for both models take the same structure as conditional cumulative distribution functions (CDFs), and hence conducting inference is similar to testing whether one CDF stochastically dominates another; We then provide an inference approach to construct valid confidence regions for β based on inverting a properly defined test statistics.

The remainder of the paper is organized as follows: In Section 2, we formally define the stationary model and derive the set of inequalities that define the sharp identified set for the parameter of interest β under stationarity assumptions. In Section 3 we replace stationarity assumption with an independence condition and derive the sharp identified set under these conditions. We also relax the independence assumption to a weaker zero conditional median assumption and derive sharp identified set under those weaker conditions. We provide sufficient conditions for point identification for both stationary and non-stationary settings in the Appendix A.1. Section 4 extends sharp bounds results to some dynamic panel data settings and to a model with time-varying factor loads. In Section 5 we construct confidence sets for the identified set both under stationarity (Model 1) and non-stationarity (Model 2) assumptions. Section 6 provides numerical evidence on the size of the identified set in some examples; and Section 7 concludes. All proofs are collected in the Appendix A.3.

2 Identification under Stationarity Assumption

n this section we study the problem of identification under conditional stationarity assumption on the disturbance terms. As a reminder, the model we are considering is of the form

$$y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it}, \text{ where } t = 1, 2.$$

Both y_{i1}^* and y_{i2}^* are only partially observed, and both ϵ_{i1} and ϵ_{i2} are unobserved. We assume below that ϵ_{i1} and ϵ_{i2} have the same distribution conditional on the vector of covariates $\mathbf{x}_i = (x_{i1}, x_{i2})$ and the fixed effect α_i. In each period, a researcher observes only (y_{it}, d_{it}, x_{it}), where $y_{it} = \max\{y_{it}^*, c_{it}\}$ and $d_{it} = 1\{y_{it}^* \geq c_{it}\}$.

Remark 2.1 *One can also assume (for example, in the case of a two sector Roy model) that the censoring variable c_{it} is generated in a way similar to y_{it}^*: $c_{it} = x_{it}'\gamma + \kappa_i + \nu_{it}$. However, unless some assumptions are made about the stochastic relationship between $(\alpha_i, \epsilon_{i1}, \epsilon_{i2})$ and $(\kappa_i, \nu_{i1}, \nu_{i2})$, imposing a linear structure on c_{it} does not help in the identification of β, as can be seen from proofs of sharp identification results (theorems 2.1 and 3.1). Therefore, we leave the censoring variable c_{it} completely unspecified.*

The question is: how do we map assumptions made on the joint distribution of $\epsilon_{i1}, \epsilon_{i2}$ conditional on $\mathbf{x}_i$ and α_i to information about the parameter β? In cross sectional models with fixed censoring at zero, Powell (1984) showed that a conditional median independence assumption made on the distribution of $\epsilon_i | x_i$ along with some full rank conditions map into point identification. In our setup, it is not easy to reach point identification without stronger assumptions. On the one hand, maintaining a conditional median independence assumption on $\epsilon_{it} | \mathbf{x}_i$ for every t will not allow us to place any meaningful bounds on β even in the absence of censoring. This is so because we do not place any restrictions on the correlation structure of vector $(\epsilon_{i1}, \epsilon_{i2})$ (see the recent work in Rosen (2012) where this point was made for panel models with no censoring). So, then, we know that with censoring, stronger assumptions are needed to obtain any non-trivial bounds on β. One of the main contributions of this paper is to show that the bounds we derive under Model 1 assumptions below are sharp, i.e., *every* parameter in the bound is one that could have generated the data under the model assumptions. For other recent work on attaining sharpness for a class of models, see Beresteanu, Molinari, and Molchanov (2011).

Assumption 1 (Model 1: Stationarity). *$\epsilon_{i1} + \alpha_i$ and $\epsilon_{i2} + \alpha_i$ have the same distribution conditional on $\mathbf{x}_i$.*

Heuristically, the change in the conditional distribution of outcomes from period 1 to period 2 is only due to the change in the values of the regressors, and so we use this variation to garner information about β. As discussed in Arellano and Honoré (2001), the strict stationarity assumption generalizes the conditional exchangeability assumption in Honoré (1992) which itself is more general than an *i.i.d.* assumption. Obviously, the censoring complicates the problem and so below, we provide the information that the *observed data* contains about β *under Model 1.*

Next, we define the following variables[6]:

$$\begin{aligned} y_{it}^U &= y_{it}, \\ y_{it}^L &= d_{it} y_{it} + (1 - d_{it})(-\infty) \end{aligned}$$

These (observed) variables y_{it}^L and y_{it}^U constitute natural lower and upper bounds on y_{it}^*, so that we always have

$$y_{it}^L \le y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it} \le y_{it}^U \tag{2.1}$$

Note that conditional on $\mathbf{x}_i = (x_{i1}, x_{i2})$, and given Model 1 above, the random variables $\alpha_i + \epsilon_{i1}$ and $\alpha_i + \epsilon_{i2}$ have the same distribution. We have then that

$$P\{\epsilon_{i1} + \alpha_i \le \tau | \mathbf{x}_i\} = P\{\epsilon_{i2} + \alpha_i \le \tau | \mathbf{x}_i\} \qquad \forall \tau$$

Therefore, the inequalities in (2.1) naturally imply that the parameter β satisfies the following set of *conditional moment inequalities* for all values of τ and $\mathbf{x}_i$:

$$\begin{aligned} &P\{y_{i1}^U - x_{i1}'\beta \le \tau | \mathbf{x}_i\} \le P\{y_{i2}^L - x_{i2}'\beta \le \tau | \mathbf{x}_i\} \\ &P\{y_{i2}^U - x_{i2}'\beta \le \tau | \mathbf{x}_i\} \le P\{y_{i1}^L - x_{i1}'\beta \le \tau | \mathbf{x}_i\} \end{aligned} \tag{2.2}$$

We define the *identified set* B_I as

$$B_I = \{b \in B : \text{ for every } \tau \in R \text{ and } \mathbf{x}_i, \text{(2.2) holds with } \beta = b\} \tag{2.3}$$

What is crucial in studying identification of finite dimensional parameters in a model such as the one above is that the conjectured identified set be shown to be the *tightest* possible set. Heuristically, this entails showing that for *every parameter* in the identified set, there exists a model obeying Model 1 assumptions above, that can generate the *observed data*. Before we formally state the sharp identification result for this model, we want to introduce another assumption.

[6] If a lower bound on the support of y_{it}^* is finite and known, one can replace $-\infty$ in the expression for y_{it}^L with this bound.

Assumption 2 (Continuous Distribution). *$\epsilon_{it} + \alpha_i$ is continuously distributed conditional on $\mathbf{x}_i$.*

Assumption 2 is not a necessary condition for the sharpness of the identified set defined in (2.3). Rather, we use it to show that placing continuity restrictions on the error terms $\epsilon_{it} + \alpha_i$ does not help to shrink the identified set. This condition implies that the conditional moments $P\{y_{it}^L - x_{it}'\beta \leq \tau | \mathbf{x}_i\}$ and $P\{y_{it}^U - x_{it}'\beta \leq \tau | \mathbf{x}_i\}$ that enter in (2.2) are continuous functions of τ for all $\mathbf{x}_i$ in the support.

Theorem 2.1 (Stationary Model). *Under Assumptions 1 and 2, any $b \in B_I$ defined in (2.3) is observationally equivalent to β and so B_I is the sharp set.*

Remark 2.2 *The set B_I above is non empty since under the correct specification the true parameter β belongs to the set. Also, the stationarity assumptions (although restrictive) does allow for correlation between ϵ_1 and ϵ_2, and, more importantly, also allows for cross sectional heteroskedasticity.*

We want to note that the arguments assume very little between the relationship between c_{it}, x_{it}, and ϵ_{it}. Notably we allow the censoring variable to be correlated with x_{it} and ϵ_{it}. This is why we refer to this setup as **endogenous censoring**. This is in contrast to the procedure proposed in Honoré, Khan, and Powell (2002). Naturally, we also allow fixed and independent censoring as special cases.

Remark 2.3 *The sharp identification result of Theorem 2.1 can be extended to a case $T \geq 2$. Having more than two time periods will add more conditional moment inequalities to (2.2), which should in general shrink the identified set. For example, if $T = 3$, the identified set is given by*

$$P\{y_{it}^U - x_{it}'\beta \leq \tau | \mathbf{x}_i\} \leq P\{y_{is}^L - x_{is}'\beta \leq \tau | \mathbf{x}_i\} \text{ for } t, s \in \{1, 2, 3\}$$

That is, if $T = 3$ we will have six conditional moment inequalities instead of only two for $T = 2$.

An immediate corollary to Theorem 2.1 follows.

Corollary 2.1 *In addition, the model contains no information on the coefficients of time invariant regressors (i.e. regressors such that $x_{i1} = x_{i2}$).*

This is immediate since if $x_{i1} = x_{i2}$, then *for every* b in the parameter space, b also belongs to B_I since it will obey the inequalities above (so, parameters for time invariant regressors can be "set" to zero). Finally, in Appendix A.1 we provide sufficient conditions for point identification in a stationary model.

As we conclude this section, we note that one drawback of the approach discussed here is the stationarity condition. As discussed in Chen and Khan (2008), this condition rules out models with with time varying heteroskedasticity, and does not allow for time varying factor loads. In the next section we relax the stationarity assumption in Model 1 above, and replace it with an independence assumption that allows for a wider range of dependence between ϵ_1 and ϵ_2.

3 Identification under Non-Stationarity Assumption

Most of the existing work in the literature on nonstationary nonlinear panel data models requires a large number of time periods- see e.g. Moon and Phillips (2000). One exception is Chen and Khan (2008), who assumed correlated random effects. Here we look for assumptions motivated from the previous literature that aim at relaxing stationarity. The issue is that standard mean and median independence assumptions on the marginal distributions of ϵ_{it}'s do not allow us to provide *any* restrictions on β, i.e. , the *sharp* set is the *trivial* set (the original parameter space itself). The intuition is that the marginal median independence assumption places no restriction on the conditional median of $(\epsilon_{i1} - \epsilon_{i2})$. Also, mean independence assumptions do not provide any identifying power with censored data without support restrictions. So, in this paper, we relax stationarity but impose statistical independence as in Model 2 below:

Assumption 3 (Model 2: Non-Stationary). *Vector* $(\epsilon_{i1}, \epsilon_{i2})$ *is independent of* $\mathbf{x}_i$.

Notice that here, the fixed effects does not enter the above formulation and so the distribution of α_i is left completely unspecified. In addition, the random variables ϵ_{i1} and ϵ_{i2} are assumed to be jointly independent of the regressors. However, since the distribution of fixed effects α_i's is left unspecified, the identifying power of the assumption in Model 2 is equivalent to the identifying power of Model 2′ below:

Assumption 4 (Model 2′: Non-Stationary). *The difference $\Delta\epsilon_i = \epsilon_{i2} - \epsilon_{i1}$ is independent of $\mathbf{x}_i = (x_{i1}, x_{i2})$.*

Note that Model 2 assumption does not require the errors to be distributed independently of fixed effects α_i's. As before, we impose no structure on variables c_{it}, thus allowing for censoring to be to correlated with regressors and outcomes. This handles both randomly endogenous censoring and fixed censoring as special cases.

We start with constructing a sharp identified set for β. Assumptions 3 or 4 do not impose any restrictions on the distribution of fixed effects α_i's, and so we have to difference out α_i's and get the following inequalities (here we use the notation from the previous section):

$$y_{i2}^L - y_{i1}^U \leq \Delta x_i'\beta + \Delta\epsilon_i \leq y_{i2}^U - y_{i1}^L$$

where $\Delta x_i = x_{i2} - x_{i1}$ and $\Delta\epsilon_i = \epsilon_{i2} - \epsilon_{i1}$. Since we assume that $\boldsymbol{\epsilon}_i$ is independent of $\mathbf{x}_i$ this means that $\Delta\epsilon$ is independent of $\mathbf{x}_i$. This will allow us to place inequality restrictions on distributions. The following theorem characterizes the sharp identified set for β under Model 2′ above.

Theorem 3.1 (Non-Stationary Model). *For any b in the parameter set B, define*

$$LB(\tau, \mathbf{x}_i, b) = P\{y_{i2}^U - y_{i1}^L - \triangle x_i'b \leq \tau | \mathbf{x}_i\}$$

and

$$UB(\tau, \mathbf{x}_j, b) = P\{y_{j2}^L - y_{j1}^U - \triangle x_j'b \leq \tau | \mathbf{x}_j\}$$

Then under Assumptions 3 and 2, the set

$$B_I = \{b \in B : \text{ for all } \mathbf{x}_i, \mathbf{x}_j \text{ and } \tau \; LB(\tau, \mathbf{x}_i, b) \leq UB(\tau, \mathbf{x}_j, b)\} \tag{3.1}$$

is the sharp identified set for β.

Similar to the stationary model, having more time periods in general will result in a smaller identified set, as more conditional moment inequalities will enter the definition of B_I.

The inequalities in (3.1) must hold *for all pairs* $(\mathbf{x}_i, \mathbf{x}_j)$. That is, slope coefficients for regressors that do not change over time cannot be separated from the fixed effects α_i's, and therefore cannot be identified.

The size of the identified set B_I also depends on the proportion of observations that are censored. If $d_{it} \equiv 1$ for all i and t, i.e. no censoring occurs, then $B_I = \{\beta\}$, i.e. parameter β is point identified. However, for the identification to be trivial, i.e. the model contains no information about β, one does not require $d_{it} \equiv 0$ for all i and t. The following result shows that in certain cases of heavy censoring, the identified set B_I coincides with the parameter space B, and so the bounds are the trivial ones.

Theorem 3.2 (Heavy Censoring). *For $t = 1,2$ denote the probability of censoring by $p_t^c(\mathbf{x}_i) = 1 - P(d_{it} = 1|\mathbf{x}_i) = P\{y_{it} < c_{it}|\mathbf{x}_i\}$. If for all $\mathbf{x}_i$ and $\mathbf{x}_j$ we have $p_1^c(\mathbf{x}_i) + p_2^c(\mathbf{x}_j) \geq 1$, then any $b \in B$ is observationally equivalent to β, so that $B_I = B$.*

This result basically says that even under the independence assumption, Model 2 (non-stationarity) provides no information on the parameter of interest, β, if there is a lot of censoring. For example, if for each $\mathbf{x}_i$ in the support at least 50% of observations are censored, then we cannot learn anything about β under the assumptions of Model 2 without making some additional assumptions.

As in the previous section, we provide next sufficient conditions for the β to be *point identified* (see Appendix).

It might be useful to characterize the above identified set as the optimizer of some objective function, and hence express the above problem as an M or U-estimation problem with a possibly non-unique optimum. It turns out that this identified set B_I can also be characterized as a set of zeros (or the Argmin set, which is the same in this case) of a particularly defined objective function. For instance, let τ_{1i}, τ_{2i} be two *i.i.d.* random variables that are continuously distributed on $(-\infty, +\infty)$ and that are independent of $\mathbf{x}_i, \mathbf{x}_j$. Let $w_i^L = y_{i2}^L - y_{i1}^U$ and $w_i^U = y_{i2}^U - y_{i1}^L$. Also, let $\tau_i = (\tau_{i1}, \tau_{i2})$. For any $b \in B$, define

$$Q(b) = E_{\tau,x}\left[1\{\tau_{2j} - \triangle x_j' b \geq \tau_{1i} - \triangle x_i' b\} 1\{P\{w_i^U \leq \tau_{1i}|\mathbf{x}_i, \tau_i\} > P\{w_j^L \leq \tau_{2j}|\mathbf{x}_j, \tau_i\}\}\right]$$

The following result shows that the identified set B_I defined above can be characterized as the set of zeros or the Argmin set of function $Q(b)$.

Theorem 3.3 (Sharp Set and Optimization Problem). *Assume that random variables τ_{1i} and τ_{2i} are independent and identically continuously distributed with support $(-\infty, +\infty)$. Further, assume that τ_{1i} and τ_{2i} are independent of x_i and x_j. Let $B_Q = \{b : Q(b) = 0\}$. Then $B_I = B_Q = \operatorname{Arg}\min_{b \in B} Q(b)$.*

The above objective function is rank based, but in the case where the regressors have continuous support, the function contains conditional probabilities *inside* indicator functions, and these conditional probabilities need to be estimated nonparametrically in a first step as was done in Khan and Tamer (2009). Note also, that the objective function $Q(\cdot)$ defined above will admit a *unique* minimum under the conditions of Theorem A.2 (sufficient conditions for point identification) in the Appendix. So, maintaining these sufficient point identification conditions, one is able to obtain a consistent estimator of β by taking the argmin of an appropriate sample analogue of $Q(\cdot)$.

3.1 Zero Conditional Median Model

Note that in the preceding discussion of the identification under non-stationarity we did not restrict the relationship between transitory error terms $(\epsilon_{i1}, \epsilon_{i2})$ and fixed effects α_i's. Therefore, the key identifying assumption is that the vector of error terms $(\epsilon_{i1}, \epsilon_{i2})$ is statistically independent of the vector of regressors x_i can be relaxed, without any loss of the identifying power, to the assumption that only the difference $\Delta\epsilon_i = \epsilon_{i2} - \epsilon_{i1}$ is independent of x_i. In this subsection, we further relax the statistical independence assumption and consider identification under the median independence assumption on the *difference in the errors*. That is, we assume that $\text{Med}(\Delta\epsilon_i|x_i) = 0$. In this case the identified set is also characterized by a set of conditional inequalities.

Assumption 5 (Model 3: Zero Conditional Median). $\text{Med}(\Delta\epsilon_i|\mathbf{x}_i) = 0$.

Theorem 3.4 (Zero Conditional Median Model). *Suppose that Assumption 5 holds. Then a sharp identified set B_I is given by $B_I = \{b \in B :$ for every $\mathbf{x}_i, \mathbf{x}_j$ $\text{Med}(y^L_{i2} - y^U_{i1}|\mathbf{x}_i) - \Delta x_i' b \leq 0 \leq \text{Med}(y^U_{j2} - y^L_{j1}|\mathbf{x}_j) - \Delta x_j' b\}$.*

Assumption 5 (Model 3) is not easy to characterize in terms of restrictions on the correlation between idiosyncratic error terms ϵ_1 and ϵ_2. For example, if ϵ_1 is independent and identically distributed to ϵ_2 (conditional on $\mathbf{x}$'s), then their difference is distributed symmetrically around 0. However, this is not the necessary condition for the zero conditional median assumption. Another example would be when ϵ_1 has a symmetric around 0 distribution, while $\epsilon_2 = \rho\epsilon_1 + \kappa$, where κ is independent of ϵ_1 and also is symmetrically distributed around 0.

4 Extensions

In this section we consider extension of our results for various dynamic panel data models, and a model with time varying factor loads. In particular, we look at what can be learned in some dynamic panel data models under the assumption of stationarity (Model 1). Also, we show that a time varying factor loads model implies non-stationarity (Model 2).

4.1 Dynamic Panel Data Models

One of the limitations of the models considered in the previous sections was the strict exogeneity condition imposed on the explanatory variables. This assumption rules out any type of dynamic feedback, such as including lagged dependent variable as an explanatory variable. Although there is much progress in dynamic linear panel data models, see Hsiao (1986), Baltagi (1995), and especially Arellano and Bond (1991), there are very few results for censored models like those considered here. Honoré (1993), Honoré and Hu (2004), and Hu (2002) all provide results for panel data dynamics with *fixed censoring*, while none of these allow for the random, endogenous censoring considered here, nor do they attain the sharp bounds when point identification is not attainable. Consequently, in this section we will consider dynamic censored panel data models (with no restrictions on censoring variable). We allow the dynamic feedback to enter in three different ways: through lagged observable variables (either the (potentially censored) outcome itself or the censoring indicator) or through the lagged latent outcome (which is only partially observable). The approach is very similar to the static censored panel data models treated in previous sections, so for the purpose of illustration we focus mostly on a stationary setting. In the non-stationarity setting, the only substantial difference between the static and the dynamic setup is for the model with lagged latent outcome, so we treat this case separately.

4.1.1 Lagged Observed Outcome

The first dynamic panel model we consider is with a lagged *observed* outcome as one of the explanatory variables:

$$y_{it}^* = \gamma y_{i,t-1} + x_{it}'\beta + \alpha_i + \epsilon_{it} \tag{4.1}$$

where $y_{it} = \max\{y_{it}^*, c_{it}\}$ so, unlike y_{it}^*, it is observed. The parameters of interest are γ and β, and in this section we will impose a conditional stationarity assumption on the disturbance terms ϵ_{it}. The autoregressive parameter γ is a determinant of the persistence of the process and is often the object of interest in empirical applications. For example, y_{it}^* is current wage in sector 1 in a two sector economy, and y_{it-1} is last period's observed wage (regardless whether individual i was employed in sector 1 or 2).

In the dynamic setting, stationarity condition on the error terms translates into the following assumption.

Assumption 6 (Stationarity). *Error terms $\epsilon_{i1} + \alpha_i$ and $\epsilon_{i2} + \alpha_i$ are identically distributed conditional on (x_{i1}, x_{i2}, y_{i0}).*

Assumption 6 is similar to the stationarity condition used in e.g. Hu (2002) for a dynamic censored panel data model.

Using the notation introduced previously, we construct conditional moments inequalities that place restrictions on β and γ as follows:

$$\begin{aligned} P\{y_{i1}^U - gy_{i0} - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}\} \leq P\{y_{i2}^L - gy_{i1} - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}\} \\ P\{y_{i2}^U - gy_{i1} - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}\} \leq P\{y_{i1}^L - gy_{i0} - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}\} \end{aligned} \tag{4.2}$$

Theorem 4.1 (Dynamic Model with Lagged Observed Outcome). *Suppose that Assumption 6 holds. Let $(g, b) \in \Theta = \Gamma \times B$ satisfy the inequalities in (4.2) for all $\tau \in \mathbb{R}$ and all $\mathbf{x}_i$ and y_{i0} in the support. The (g, b) is observationally equivalent to the true parameter value (γ, β). That is, the sharp identified set is*

$$\Theta_{I,1} = \{(g, b) \in \Theta : \textit{ for every } \tau \in \mathbb{R} \textit{ and every } \mathbf{x}_i \textit{ and } y_{i0} \text{ (4.2) } \textit{hold.}\}$$

4.1.2 Lagged Censoring Indicator

The second dynamic panel data model we consider is when a lagged value of the *censoring indicator* variable d_{it} as used an explanatory variable, and we maintain the initial conditions assumption as before (i.e., we observe d_{i0}). This is an interesting model where the dynamics of the outcome process is through the sector specific lagged variable:

$$y_{it}^* = \gamma d_{i,t-1} + x_{it}'\beta + \alpha_i + \epsilon_{it} \tag{4.3}$$

Specifically, for the first two periods we have:

$$\begin{aligned} y_{i1}^* &= \alpha_i + \gamma d_{i0} + x_{i1}'\beta + \epsilon_{i1} \\ y_{i2}^* &= \alpha_i + \gamma d_{i1} + x_{i2}'\beta + \epsilon_{i2} \end{aligned}$$

Here the stationarity condition is summarized by Assumption 7.

Assumption 7 (Stationarity). *Error terms $\epsilon_{i1} + \alpha_i$ and $\epsilon_{i2} + \alpha_i$ are identically distributed conditional on (x_{i1}, x_{i2}, d_{i0}).*

Assumption 7 gives us a set of conditional moment inequalities analogous to what we had before:

$$\begin{aligned} P\{y_{i1}^U - gd_{i0} - x_{i1}'b \le \tau | \mathbf{x}_i, d_{i0}\} &\le P\{y_{i2}^L - gd_{i1} - x_{i2}'b \le \tau | \mathbf{x}_i, d_{i0}\} \\ P\{y_{i2}^U - gd_{i1} - x_{i2}'b \le \tau | \mathbf{x}_i, d_{i0}\} &\le P\{y_{i1}^L - gy_{i0} - x_{i1}'b \le \tau | \mathbf{x}_i, d_{i0}\} \end{aligned} \tag{4.4}$$

The identified set for $\theta = (\gamma, \beta')'$ given by the inequalities in 4.4 is sharp according to the result below.

Theorem 4.2 (Dynamic Model with Lagged Censoring Indicator). *Suppose that Assumption 7 holds. Let $(g, b) \in \Theta = \Gamma \times B$ satisfy the inequalities in (4.4) for all $\tau \in \mathbb{R}$ and all $\mathbf{x}_i$ and y_{i0} in the support. The (g, b) is observationally equivalent to the true parameter value (γ, β). That is, the sharp identified set is*

$$\Theta_{I,2} = \{(g, b) \in \Theta : \text{ for every } \tau \in \mathbb{R} \text{ and every } \mathbf{x}_i \text{ and } d_{i0} \text{ (4.4) hold.}\}$$

4.1.3 Lagged Latent Outcome

The third model we consider is when the lagged value of the *latent* outcome y_{it}^* is one of the explanatory variables:

$$y_{it}^* = \gamma y_{i,t-1}^* + x_{it}'\beta + \alpha_i + \epsilon_{it} \tag{4.5}$$

Following Hu (2002) we assume that y_{i0}^* is observed (i.e., there is no censoring in the initial period). For the first and second period we have

$$\begin{aligned} y_{i1}^* &= \alpha_i + \gamma y_{i0}^* + x_{i1}'\beta + \epsilon_{i1} \\ y_{i2}^* &= \alpha_i + \gamma y_{i1}^* + x_{i2}'\beta + \epsilon_{i2} \end{aligned}$$

where y_{i1}^* is not necessarily observed.

Assumption 8 (Stationarity). *Error terms $\epsilon_{i1} + \alpha_i$ and $\epsilon_{i2} + \alpha_i$ are identically distributed conditional on $(x_{i1}, x_{i2}, y_{i0}^*)$.*

First, we assume that $\gamma \geq 0$ (same assumption is made in Hu (2002)). Then we have the following inequalities:

$$\begin{aligned} y_{i1}^L - \gamma y_{i0}^* - x_{i1}'\beta \leq \epsilon_{i1} + \alpha_i \leq y_{i1}^U - \gamma y_{i0}^* - x_{i1}'\beta \\ y_{i2}^L - \gamma y_{i1}^U - x_{i2}'\beta \leq \epsilon_{i2} + \alpha_i \leq y_{i2}^U - \gamma y_{i1}^L - x_{i2}'\beta \end{aligned}$$

It is easy to see that Assumption 8 implies that for every τ and $\mathbf{x}_i$, y_{i0}^*, the following inequalities must hold:

$$\begin{aligned} P\{y_{i1}^U - g y_{i0}^* - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \leq P\{y_{i2}^L - g y_{i1}^U - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \\ P\{y_{i2}^U - g y_{i1}^L - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \leq P\{y_{i1}^L - g y_{i0}^* - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \end{aligned} \tag{4.6}$$

However, unlike the static panel data model with stationarity assumption, here the inequalities in (4.6) do not give us the **sharp** identified set. On the one hand, we can show that any vector of parameters (g, b) that is observationally equivalent to true parameters (γ, β) will also satisfy the inequalities in (4.6) (with γ and β replaced by g and b, respectively). However, the upper and lower bounds in (4.6) with respect to partially observed y_{i1}^* are not independent, and therefore we cannot expect that any parameters that satisfy inequalities in (4.6) are observationally equivalent to the true parameters. However, if we consider a dynamic analog in a non-stationary setting, then with $T = 2$ periods we can characterize the identified set through a set of conditional moment inequalities as before. The set defined in (4.6) can only be shown to contain the identified set[7]. In the Appendix A.2 we characterize the sharp identified set under the stationarity assumption.

4.1.4 Lagged Latent Outcome in a Non-Stationary Model

In the dynamic analog of the non-stationary case, one can construct a set of conditional inequalities that give a sharp set when $T = 2$. Here we construct the identified set for a dynamic model with lagged latent outcome under the non-stationarity assumption of Section 3.

Suppose that we know γ. If we subtract first-period equation from the second-period equation, we have

$$y_{i2}^* - (1+\gamma)y_{i1}^* + \gamma y_{i0}^* = \Delta x_i'\beta + \epsilon_{i2} - \epsilon_{i1}$$

We make the following assumption (similar to Assumption 3) about the error terms:

Assumption 9 (Non-Stationary Dynamic Model). *Vector $(\epsilon_{i1}, \epsilon_{i2})$ is independent of $\mathbf{x}_i$.*

If $1+\gamma \geq 0$, then we can work with the following inequalities:

$$y_{i2}^L - (1+\gamma)y_{i1}^U + \gamma y_{i0}^* \leq y_{i2}^* - (1+\gamma)y_{i1}^* + \gamma y_{i0}^* \leq y_{i2}^U - (1+\gamma)y_{i1}^L + \gamma y_{i0}^*$$

Again, we start by assuming that $1+\gamma \geq 0$. Then for a candidate (g, b) we can subtract $\Delta x_i'b$ and check whether the following inequalities hold for every τ, $\mathbf{x}_i$, $\mathbf{x}_j$, y_{i0}^* and y_{j0}^*:

$$P\{y_{i2}^U - (1+g)y_{i1}^L + gy_{i0}^* - \Delta x_i'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \leq P\{y_{j2}^L - (1+g)y_{j1}^U + gy_{j0}^* - \Delta x_j'b \leq \tau | \mathbf{x}_j, y_{j0}^*\} \tag{4.7}$$

A set of parameters that satisfy (4.7) gives the identified set, as summarized below.

[7] Following Hu (2002), we can also assume conditional independence between ϵ_{i1} and ϵ_{i2}, which implies that $\Delta\epsilon_i = \epsilon_{i2} - \epsilon_{i1}$ is symmetrically distributed conditional on $\mathbf{x}_i$ and exploit additional conditional moment inequalities imposed by the symmetry. This potentially can shrink the set in (4.6), but it does not guarantee the sharpness of the set.

Theorem 4.3 (Non-Stationary Dynamic Model). *Suppose that Assumption 9 holds, and assume that $1+\gamma > 0$. Let $(g,b) \in \Theta = \Gamma \times B$, $1+g>0$ satisfy the inequalities in (4.7) for all $\tau \in \mathbb{R}$ and all $\mathbf{x}_i$, $\mathbf{x}_j$, y^*_{i0} and y^*_{j0}. Then (g,b) is observationally equivalent to the true parameter value (γ,β). That is, the sharp identified set is*

$$\Theta_{I,4} = \{(g,b) \in \Theta,\ 1+g>0: \textit{ for every } \tau \in \mathbb{R} \textit{ and every } \mathbf{x}_i,\ \mathbf{x}_j,\ y^*_{i0} \textit{ and } y^*_{j0} \text{ (4.7) } \textit{hold.}\}$$

4.2 Time Varying Factor Loads

A particular nonstationary panel data model that has received interest in empirical settings is one where a time varying factor loads onto the individual specific effect. Maintaining our notation, we can express the latent equation as:

$$y^*_{it} = \gamma_t \alpha_i + x'_{it}\beta + \epsilon_{it}$$

where γ_t denotes the time varying factor load. This parameter is of interest in labor economics as it represents the returns to unobserved skills, which may change over time (see, e.g. Chay and Honoré (1998)). We can easily modify our approach to attain bounds on β and γ_t, assuming cross sectional homoskedasticity

We illustrate the idea using two periods as we did before. Note here we can only identify the ratio $\gamma_2/\gamma_1 = \gamma$, so we normalize $\gamma_1 \equiv 1$ and $\gamma_2 = \gamma$. Then the equations for periods $t=1$ and $t=2$ become

$$\begin{aligned} y^*_{i1} &= \alpha_i + x'_{i1}\beta + \epsilon_{i1} \\ y^*_{i2} &= \gamma\alpha_i + x'_{i2}\beta + \epsilon_{i2} \end{aligned}$$

We assume that $\gamma \neq 0$ and so we can divide both sides of the second equation by γ.[8]

$$y^*_{i2}/\gamma = \alpha_i + x'_{i2}\beta/\gamma + \epsilon_{i2}/\gamma$$

This division immediately results in the nonstationarity of the error terms. However, if we assume independence (a non-stationarity assumption), we can place meaningful

[8] Here $\gamma = 0$ would imply that the fixed effect does not enter in time period 2, which does not seem plausible.

restrictions on the model's parameters β in γ. In particular, assume first that $\gamma > 0$. Then we can difference out α_i, so that the upper and lower bounds on $\epsilon_{i2}/\gamma - \epsilon_{i1}$ are

$$y_{i2}^L/\gamma - y_{i1}^U - (x_{i2}'/\gamma - x_{i1})'\beta \leq \epsilon_{i2}/\gamma - \epsilon_{i1} \leq y_{i2}^U/\gamma - y_{i1}^L - (x_{i2}'/\gamma - x_{i1})'\beta$$

Then under Assumption 3 and assuming that $\gamma > 0$, the identified set is summarized by the following conditional moment inequalities that must hold for all values of τ and all $\mathbf{x}_i$, $\mathbf{x}_j$ in the support:

$$P\{y_{i2}^U/g - y_{i1}^L - (x_{i2}/g - x_{i1})'b \leq \tau|\mathbf{x}_i\} \leq P\{y_{j2}^L/g - y_{j1}^U - (x_{j2}/g - x_{j1})'b \leq \tau|\mathbf{x}_j\} \quad (4.8)$$

If $\gamma < 0$, the identified set is given by

$$P\{y_{i2}^L/g - y_{i1}^L - (x_{i2}/g - x_{i1})'b \leq \tau|\mathbf{x}_i\} \leq P\{y_{j2}^U/g - y_{j1}^U - (x_{j2}/g - x_{j1})'b \leq \tau|\mathbf{x}_j\}$$

This is summarized by the following claim.

Theorem 4.4 (Time Varying Factor Loads). *Suppose that Assumption 3 holds, and assume that $\gamma > 0$. Let $(g, b) \in \Theta = \Gamma \times B$, $g > 0$ satisfy the inequalities in (4.8) for all $\tau \in \mathbb{R}$ and all $\mathbf{x}_i$, $\mathbf{x}_j$. Then (g, b) is observationally equivalent to the true parameter value (γ, β). That is, the sharp identified set is*

$$\Theta_{I,TVFL} = \{(g, b) \in \Theta,\ g > 0 : \textit{ for every } \tau \in \mathbb{R} \textit{ and every } \mathbf{x}_i \textit{ and } \mathbf{x}_j \text{ (4.8) } \textit{hold.}\}$$

Having more than $T = 2$ time periods adds to the number of inequalities that define the identified set. However, it adds more parameters, too. Therefore we cannot say any longer that observing more time periods helps to shrink the identified set (as we were able to say before). For example, for $T = 3$ we now have three parameters: γ_2, γ_3 and β. The equations for periods 1, 2 and 3 now are

$$\begin{aligned} y_{i1}^* &= \alpha_i + x_{i1}'\beta + \epsilon_{i1} \\ y_{i2}^* &= \gamma_2\alpha_i + x_{i2}'\beta + \epsilon_{i2} \\ y_{i3}^* &= \gamma_3\alpha_i + x_{i2}'\beta + \epsilon_{i2} \end{aligned}$$

and the identified set for $(\gamma_2, \gamma_3, \beta)$ is given by (assuming e.g. that $\gamma_2, \gamma_3 > 0$):

$$
\begin{aligned}
P\left\{\frac{y_{i2}^U}{g_2} - y_{i1}^L - \left(\frac{x_{i2}}{g_2} - x_{i1}\right)' b \le \tau|\mathbf{x}_i\right\} &\le P\left\{\frac{y_{j2}^L}{g_2} - y_{j1}^U - \left(\frac{x_{j2}}{g_2} - x_{j1}\right)' b \le \tau|\mathbf{x}_j\right\} \\
P\left\{\frac{y_{i3}^U}{g_3} - y_{i1}^L - \left(\frac{x_{i3}}{g_3} - x_{i1}\right)' b \le \tau|\mathbf{x}_i\right\} &\le P\left\{\frac{y_{j3}^L}{g_3} - y_{j1}^U - \left(\frac{x_{j3}}{g_3} - x_{j1}\right)' b \le \tau|\mathbf{x}_j\right\} \\
P\left\{\frac{y_{i3}^U}{g_3} - \frac{y_{i2}^L}{g_2} - \left(\frac{x_{i3}}{g_3} - \frac{x_{i2}}{g_2}\right)' b \le \tau|\mathbf{x}_i\right\} &\le P\left\{\frac{y_{j3}^L}{g_3} - \frac{y_{j2}^U}{g_2} - \left(\frac{x_{j3}}{g_3} - \frac{x_{j2}}{g_2}\right)' b \le \tau|\mathbf{x}_j\right\}
\end{aligned}
$$

That is, now we have three inequalities instead of only one for $T = 2$, but we now also have one additional parameter.

5 Inference

This section provides an approach for statistical inference given the identification results in previous sections. We suggest methods that can be used to build confidence regions for β, taking into account the fact that this parameter, in most of the cases above, might not be point identified. We provide assumptions under which the large sample distribution of a test statistic is derived and used to construct confidence regions on the true parameter. There has been a lot of work on the statistical inference of models that are partially identified, and so this section mostly adapts some methods from the recent literature. We focus on the case when covariates $\mathbf{x}$ have discrete distribution with a finite support[9].

We construct confidence intervals for β under stationarity or non-stationarity restrictions based on pointwise testing inequalities in (2.2) and (3.1) correspondingly. That is, the confidence set for β will collect all candidate parameter values for which one failed to reject the null hypothesis that this candidate value belongs to the identified set. In both Model 1 and 2, we use a Kolmogorov-Smirnov type test statistic that is based on

[9]The case of continuous covariates will require a non-parametric estimation of conditional moment inequalities for any fixed value of τ, which, together with continuous τ, is too involved and is beyond the scope of this paper. Some discussion of moment inequalities models with continuous covariates can be found in, for example, Andrews and Shi (2013), Chernozhukov, Lee, and Rosen (2013), Kim (2007), and Ponomareva (2010).

the corresponding conditional moment inequalities. Before we proceed, we introduce additional assumptions that will be used throughout this section:

Assumption 10 (Random Sampling). *Observations* $\{(y_{it}, d_{it}, x'_{it})'; \ t = 1, 2, \ i = 1 \dots, n\}$ *are i.i.d. across individuals.*

Assumption 11 $\inf_{\mathbf{x} \in \mathcal{X}} P\{\mathbf{x}_i = \mathbf{x}\} \geq \delta > 0$ *for some* $\delta > 0$.

Assumption 10 is a standard random sampling assumption, and Assumption 11 is a regularity condition that ensures that we can estimate conditional moments that enter (2.3) at $n^{1/2}$-rate.

5.1 Inference in the Stationary Model

According to Theorem 2.1, each β in the identified set B_I should satisfy the following inequalities for all values of τ and $\mathbf{x}_i$:

$$
\begin{aligned}
&P\{y^U_{i1} - x'_{i1}\beta \leq \tau | \mathbf{x}_i\} \leq P\{y^L_{i2} - x'_{i2}\beta \leq \tau | \mathbf{x}_i\} \\
&P\{y^U_{i2} - x'_{i2}\beta \leq \tau | \mathbf{x}_i\} \leq P\{y^L_{i1} - x'_{i1}\beta \leq \tau | \mathbf{x}_i\}
\end{aligned}
\tag{5.1}
$$

For a candidate value β, we define $D_1(\tau, \mathbf{x}_i; \beta) = P\{y^U_1 - x'_1\beta \leq \tau | \mathbf{x}\} - P\{y^L_2 - x'_2\beta \leq \tau | \mathbf{x}\}$ and $D_2(\tau, \mathbf{x}; \beta) = P\{y^U_2 - x'_2\beta \leq \tau | \mathbf{x}\} - P\{y^L_1 - x'_1\beta \leq \tau | \mathbf{x}\}$. Therefore, testing whether or not a candidate parameter value β belongs to the identified set amounts to testing the following null vs alternative hypotheses:

$$
\begin{aligned}
&H_{S,0} : D_1(\tau, \mathbf{x}; \beta) \leq 0, \text{ and } D_2(\tau, \mathbf{x}; \beta) \leq 0 \text{ for all } (\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X} \\
&H_{S,a} : \max\{D_1(\tau, \mathbf{x}; \beta), D_2(\tau, \mathbf{x}; \beta)\} > 0 \text{ for some } (\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X}
\end{aligned}
\tag{5.2}
$$

Assume for a moment that we have a test of size α of H_0 against H_a for each value of β. Then the $(1 - \alpha) \cdot 100\%$ confidence set for β will collect all values of beta for which we failed to reject H_0. Below we describe a test of $H_{S,0}$ against $H_{S,a}$.

The sample analogs of $D_1(\tau, \mathbf{x}; \beta)$ and $D_2(\tau, \mathbf{x}; \beta)$ are:

$$\hat{D}_1(\tau, \mathbf{x}; \beta) = \frac{1}{n}\sum_{i=1}^{n}\Big[1\{y_{i1}^U - x_{i1}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i2}^L - x_{i2}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\}\Big] / \hat{P}\{\mathbf{x}_i = \mathbf{x}\}$$

$$\hat{D}_2(\tau, \mathbf{x}; \beta) = \frac{1}{n}\sum_{i=1}^{n}\Big[1\{y_{i2}^U - x_{i2}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i1}^L - x_{i1}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\}\Big] / \hat{P}\{\mathbf{x}_i = \mathbf{x}\}$$

where $\hat{P}\{\mathbf{x}_i = \mathbf{x}\} = \frac{1}{n}\sum_{i=1}^{n} 1\{\mathbf{x}_i = \mathbf{x}\}$

To test the null hypothesis in (5.2) we use the following Kolmogorov-Smirnov type test statistic:

$$T_n^S(\beta) = \max\{\sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} \sqrt{n}\hat{D}_1(\tau, \mathbf{x}; \beta), \sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} \sqrt{n}\hat{D}_2(\tau, \mathbf{x}; \beta)\}$$

The limiting distribution of $T_n^S(\beta)$ under the null hypothesis is given in the result below.

Theorem 5.1 (Distribution of Test Statistic in a Stationary Model). *Suppose that Assumptions 1,2, 10 and 11 hold. Then under $H_{S,0}$, the limiting distribution of $T_n^S(\beta)$ is first-order stochastically dominated by the distribution of*

$$\max\{\sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} G_1(\tau, \mathbf{x}; \beta), \sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} G_2(\tau, \mathbf{x}; \beta)\}$$

where $G^S(\tau, \mathbf{x}; \beta) = (G_1(\tau, \mathbf{x}; \beta), G_2(\tau, \mathbf{x}; \beta))$ is a two-dimensional gaussian process with zero mean and continuous sample paths in $\ell^\infty(\mathbb{R}, \mathcal{X})$ and the following covariance kernel:

$$\begin{aligned}&\mathrm{Cov}(G^S(\tau_1,\mathbf{x}_1;\beta), G^S(\tau_2,\mathbf{x}_2;\beta)) = \\ &\quad = \begin{pmatrix}\mathrm{Cov}(\delta_{i1}(\tau_1,\mathbf{x}_1;\beta), \delta_{i1}(\tau_2,\mathbf{x}_2;\beta)) & \mathrm{Cov}(\delta_{i1}(\tau_1,\mathbf{x}_1;\beta), \delta_{i2}(\tau_2,\mathbf{x}_2;\beta)) \\ \mathrm{Cov}(\delta_{i2}(\tau_1,\mathbf{x}_1;\beta), \delta_{i1}(\tau_2,\mathbf{x}_2;\beta)) & \mathrm{Cov}(\delta_{i2}(\tau_1,\mathbf{x}_1;\beta), \delta_{i2}(\tau_2,\mathbf{x}_2;\beta))\end{pmatrix}\end{aligned}$$

where

$$\delta_{i1}(\tau, \mathbf{x}; \beta) = \frac{1\{y_{i1}^U - x_{i1}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i2}^L - x_{i2}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\}}{P\{\mathbf{x}_i = \mathbf{x}\}}$$

$$\delta_{i2}(\tau, \mathbf{x}; \beta) = \frac{1\{y_{i2}^U - x_{i2}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i1}^L - x_{i1}'\beta \le \tau, \mathbf{x}_i = \mathbf{x}\}}{P\{\mathbf{x}_i = \mathbf{x}\}}$$

We can now use the limiting distribution of $T_n^S(\beta)$ to construct confidence set for the parameter of interest, β.

Corollary 5.1 (Confidence Set in a Stationary Model). *Suppose that assumptions of Theorem 5.1 hold. Let $c^S_{1-\alpha}(\beta)$ denote the $(1-\alpha)\cdot 100\%$ quantile of the distribution of*

$$\max\{\sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} G_1(\tau,\mathbf{x};\beta), \sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} G_2(\tau,\mathbf{x};\beta)\}$$

and define the $(1-\alpha)\cdot 100\%$ confidence set for β as

$$CS_{n,1-\alpha} = \{\beta \in B : T^S_n(\beta) \leq c^S_{1-\alpha}(\beta)\}$$

Then $\lim_{n\to\infty} P\{\beta \in CS_{n,1-\alpha}\} \geq 1-\alpha$.

Since $P\{\mathbf{x}_i = \mathbf{x}\}$ needs to be estimated, we have to estimate the covariance structure of the gaussian process $G^S(\tau,\mathbf{x};\beta)$, and then we can sample from that estimated process. We can consistently estimate the above covariance structure using sample covariances, and replacing the unknown $P\{\mathbf{x}_i = \mathbf{x}\}$ with its consistent estimator $\hat{P}\{\mathbf{x}_i = \mathbf{x}\}$. Finally, to estimate critical values $c^S_{1-\alpha}(\beta)$, we can approximate the gaussian process $(G_1(\tau,\mathbf{x};\beta), G_2(\tau,\mathbf{x};\beta)$ with $\hat{G}_1(\tau,\mathbf{x};\beta,Z), \hat{G}_2(\tau,\mathbf{x};\beta,Z))$ defined by

$$\hat{G}_1(\tau,\mathbf{x};\beta,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\delta}_{i1}(\tau,\mathbf{x};\beta) - \hat{D}_1(\tau,\mathbf{x};\beta))z_i$$
$$\hat{G}_2(\tau,\mathbf{x};\beta,Z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\hat{\delta}_{i2}(\tau,\mathbf{x};\beta) - \hat{D}_2(\tau,\mathbf{x};\beta))z_i$$

where $Z = (z_1,\ldots,z_n)$ and $\{z_1,\ldots,z_n\}$ are *i.i.d.* draws from a standard normal distribution, and

$$\hat{\delta}_{i1}(\tau,\mathbf{x};\beta) = \frac{1\{y^U_{i1} - x'_{i1}\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y^L_{i2} - x'_{i2}\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\}}{\hat{P}\{\mathbf{x}_i = \mathbf{x}\}}$$
$$\hat{\delta}_{i2}(\tau,\mathbf{x};\beta) = \frac{1\{y^U_{i2} - x'_{i2}\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y^L_{i1} - x'_{i1}\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\}}{\hat{P}\{\mathbf{x}_i = \mathbf{x}\}}$$

Finally, similar to Jun, Lee, and Shin (2011), we can estimate $c^S_{1-\alpha}(\beta)$ using the $(1-\alpha)\cdot 100\%$ quantile of the empirical distribution of

$$\max\{\sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} \hat{G}_1(\tau,\mathbf{x};\beta,Z), \sup_{(\tau,\mathbf{x})\in\mathbb{R}\times\mathcal{X}} \hat{G}_2(\tau,\mathbf{x};\beta,Z)\}$$

The inference procedure described above may result in the conservative coverage due to using the distribution that first-order dominates the limiting distribution of $T^S_n(\beta)$.

To improve the coverage, one can follow Linton, Song, and Whang (2010) and take the supremum of the above gaussian process over "contact sets" rather than the whole support for $\mathbf{x}$ and τ. A contact sets is defined as the set of all values $(\tau, \mathbf{x})$ that set inequalities in (5.1) to equalities. In particular, we can define the contact sets as

$$B_1(\beta) = \{(\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X} : D_1(\tau, \mathbf{x}; \beta) = 0\}$$
$$B_2(\beta) = \{(\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X} : D_2(\tau, \mathbf{x}; \beta) = 0\}$$

Then under the condition of Theorem 5.1, we have

$$T_n^S(\beta) \Rightarrow \max\{ \sup_{(\tau,\mathbf{x}) \in B_1(\beta)} G_1(\tau, \mathbf{x}; \beta), \sup_{(\tau,\mathbf{x}) \in B_2(\beta)} G_2(\tau, \mathbf{x}; \beta)\}$$

Those contact sets can be consistently estimated by

$$\hat{B}_1(\beta) = \left\{(\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X} : |\hat{D}_1(\tau, \mathbf{x}; \beta)| \leq \frac{a_n}{\sqrt{n}}\right\}$$
$$\hat{B}_2(\beta) = \left\{(\tau, \mathbf{x}) \in \mathbb{R} \times \mathcal{X} : |\hat{D}_2(\tau, \mathbf{x}; \beta)| \leq \frac{a_n}{\sqrt{n}}\right\}$$

where $\{a_n > 0 : n = 1, 2, \ldots\}$ is a deterministic sequence such that $a_n \to 0$, $a_n/\sqrt{n} \to 0$ and $\sqrt{\ln \ln n}/a_n \to 0$. Though it is feasible to estimate these contact sets above, using the distribution where we take the supremum over all $(\tau, \mathbf{x})$ is easier to construct, especially in the case where $\mathbf{x}$ has discrete support.

5.2 Inference in the Non-Stationary Model

For a non-stationary model, the identified set for β is given by the following set of inequalities (see Theorem 3.1):

$$P\{y_{i2}^U - y_{i1}^L - \Delta x_i' b \leq \tau | \mathbf{x}_i\} \leq P\{y_{j2}^L - y_{j1}^U - \Delta x_j' b \leq \tau | \mathbf{x}_j\} \tag{5.3}$$

For a fixed candidate β, we define

$$D(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta) = P\{y_{i2}^U - y_{i1}^L - \Delta x_i' \beta \leq \tau | \mathbf{x}_i = \mathbf{x}\} - P\{y_{j2}^L - y_{j1}^U - \Delta x_j' \beta \leq \tau | \mathbf{x}_j = \tilde{\mathbf{x}}\}$$

Then the null and the alternative hypotheses for testing that this candidate β belongs to the identified set B_I in this case can be stated as

$$\begin{aligned} &H_{NS,0} : D(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta) \leq 0 \text{ for all } (\tau, \mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X} \\ &H_{NS,a} : D(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta) > 0 \text{ for some } (\tau, \mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X} \end{aligned} \tag{5.4}$$

$D(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta)$ can be consistently estimated with

$$\hat{D}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta) = \frac{1}{n} \sum_{i=1}^{n} 1\{y_{i2}^{U} - y_{i1}^{L} - \Delta x_i' \beta \le \tau, \mathbf{x}_i = \mathbf{x}\} / \hat{P}\{\mathbf{x}_i = \mathbf{x}\}$$
$$- \frac{1}{n} \sum_{j=1}^{n} 1\{y_{j2}^{L} - y_{j1}^{U} - \Delta x_j' \beta \le \tau, \mathbf{x}_j = \tilde{\mathbf{x}}\} / \hat{P}\{\mathbf{x}_j = \tilde{\mathbf{x}}\}$$

To test the null hypothesis we again use the Kolmogorov-Smirnov type test statistic:

$$T_N^{NS}(\beta) = \sup_{(\tau, \mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X}} \sqrt{n} \hat{D}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta)$$

The limiting distribution of this test statistic is summarized in the result below.

Theorem 5.2 (Distribution of Test Statistic in a Non-Stationary Model). *Suppose that Assumptions 3(4),2, 10 and 11 hold. Then under $H_{NS,0}$, the limiting distribution of $T_n^{NS}(\beta)$ is first-order stochastically dominated by the distribution of*

$$\sup_{(\tau, \mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X}} G^{NS}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta)$$

where $G^{NS}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta)$ is a gaussian process with zero mean and continuous sample paths in $\ell^{\infty}(\mathbb{R}, \mathcal{X}, \mathcal{X})$ and the following covariance kernel:

$$\text{Cov}(G^{NS}(\tau_1, \mathbf{x}_1, \tilde{\mathbf{x}}_1; \beta), G^{NS}(\tau_2, \mathbf{x}_2, \tilde{\mathbf{x}}_2; \beta)) = \text{Cov}\left(\delta_{ij}(\tau_1, \mathbf{x}_1, \tilde{\mathbf{x}}_1; \beta), \delta_{ij}(\tau_2, \mathbf{x}_2, \tilde{\mathbf{x}}_2; \beta)\right)$$

where

$$\delta_{ij}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta) = \frac{1\{y_{i2}^{U} - y_{i1}^{L} - \Delta x_i' \beta \le \tau, \mathbf{x}_i = \mathbf{x}\}}{P\{\mathbf{x}_i = \mathbf{x}\}} - \frac{1\{y_{j2}^{L} - y_{j1}^{U} - \Delta x_j' \beta \le \tau, \mathbf{x}_j = \tilde{\mathbf{x}}\}}{P\{\mathbf{x}_j = \tilde{\mathbf{x}}\}}$$

We use the limiting distribution of $T_n^{NS}(\beta)$ to construct confidence set for the parameter of interest, β.

Corollary 5.2 (Confidence Set in a Non-Stationary Model). *Suppose that assumptions of Theorem 5.2 hold. Let $c^{NS}1 - \alpha(\beta)$ denote the $(1-\alpha) \cdot 100\%$ quantile of the distribution of*

$$\sup_{(\tau, \mathbf{x}, \tilde{\mathbf{x}}) \in \mathbb{R} \times \mathcal{X} \times \mathcal{X}} G^{NS}(\tau, \mathbf{x}, \tilde{\mathbf{x}}; \beta)$$

and define the $(1-\alpha) \cdot 100\%$ confidence set for β as

$$CS_{n,1-\alpha} = \{\beta \in B : T_n^{NS}(\beta) \le c_{1-\alpha}^{NS}(\beta)\}$$

Then $\lim_{n \to \infty} P\{\beta \in CS_{n,1-\alpha}\} \ge 1 - \alpha$.

As in the stationary case, one can approximate the gaussian process $G^{NS}(\cdot,\cdot,\cdot;\beta)$ and sample from this approximation to estimate critical values $c^{NS}1-\alpha(\beta)$. We can also improve the coverage by taking the supremum over estimated contact sets instead of the whole support.

6 Simulations

This section provides evidence on the size of the identified sets in some stylized panel models with censoring. The first set of simulations are meant to shed light on the size of the identified set in some examples, without issues of sample uncertainty (done with a very large sample size). Second, we provide small sample evidence using our inference approach to construct confidence regions on the parameters in some models also in the following section.

All the simulations (for the identified sets and the confidence regions) are based on the two period model and its dynamic variant:

$$y_t^* = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \quad t = 1, 2 \tag{6.1}$$

where $\beta_1 = \beta_0 = 1$. We use two regressors both with a discrete distribution with support on $\{-1, 0, 1\}$ in the non-dynamic models, and in the dynamic models, we only have one regressor (in addition to the lagged variables).

6.1 Identified Set Simulations

The sets of simulations here are meant to provide evidence on *the size of the identified set* in certain stylized designs. The identified set simulations are useful in their own rights: 1) for the simple models we simulate with random censoring and under various assumptions, it is not known whether the model is point identified, and 2) in many cases with endogenous censoring and/or heteroskedasticity, and though the model is not likely to be point identified, the identified sets are tight in our designs which is suggestive that under weaker conditions, these models do contain information. So, our approach then allows us to examine the sensitivity of our model to these strong assumptions. We first simulate various versions of the above under **Model 1** and **Model 2**. We start with **Model 1**.

6.1.1 Identified Set in Model 1

For this model, we plot the set of parameters (b_1, b_2) that satisfy the inequalities in (2.2). These inequalities were simulated with a sample of size 20000 for each x value (a total sample size of 16*20000) to minimize the issues of sampling uncertainty. We plot the identified set as contour plots where we use a grid point to look for parameters that do not violate any of the inequalities. For τ, we use a grid on $[-20, 20]$ with various grid sizes. Throughout, the fixed effect was generated as $\alpha_i = \mathcal{N}(0,1) * (\sum_{t=1,2;k=1,2} x_{kt})$. We start in Figure 1(a) with the panel data with fixed censoring at zero. Here, ϵ_1 is normal with mean zero and variance 2, and similarly to ϵ_2. The two random variables ϵ_1 and ϵ_2 are correlated with correlation coefficient of $1/2$. This case obeys the assumptions of Honoré (1992) and hence we expect this to be point identified and this is confirmed in the top panel of Figure 1. The second Figure, we plot the identified set also for the case with independent random censoring in which c is $\mathcal{N}(0, .25)$. The identified set here appears to be tight. For both of these designs, the level of censoring was around 30%. In the bottom panel of Figure 1, we plot the identified set for the random endogenous censoring in which $c \sim \mathcal{N}(0,1) + .5\epsilon^2$. Here, we see that the identified set is larger. There also, we plot the case with covariate dependent censoring that does not depend on ϵ. Here, $c_1 \sim N(0,1) + (x_{21} - x_{11})$ and as we can see, the identified set is smaller than the case with endogeneity. Figure 2 provides the identified set for the case with covariate dependent endogenous censoring and the bottom panel graphs the case for fixed censoring at zero where the density of ϵ is heteroskedastic. Also, we have heteroskedasticity and endogenous censoring, while in the last graph in Figure 2, we allow the censoring to depend on the covariates. Note that the largest identified sets in these designs seem to be in models with endogenous censoring, and that having the censoring depend on x in our design reduces the size of the identified set.

6.2 Identified Set in Model 2:

This is the independent non-stationary model. So, we simulate ϵ_1 as a random normal, and $\epsilon_2 \sim u \times \epsilon_1 + \frac{1}{2}z$ where u is a uniform random variable on $[-1, 1]$, and z is a standard normal independent of u and ϵ_1. On the top of Figure 3, we plot the identified set for the fixed censoring case where we have 30% censoring in period 1 and 15% in period 2. Next, we simulate the same model but with random independent censoring that is $\mathcal{N}(-\frac{1}{2}, 1)$ in period 1 and $\mathcal{N}(-1, 1)$ in period 2 which resulted in 40% and 26% censoring in periods 1 and 2 respectively. As we can see, in this design, the random censoring shrinks somehow the identified set. In the bottom of Figure 3, we have design with endogenous random censoring where the censoring in period 1 is $c_1 = \mathcal{N}(0, 1) + 2\epsilon_2 + .5$ while in period 2 it is $c_2 = \mathcal{N}(0, 1) - .1\epsilon_1 + 1$ which got us around 20% censoring in period 1 and 15% censoring in period 2. The last graph in Figure 3 provides a case where the censoring in addition to being endogenous, is also covariate dependent. Here, the censoring in both periods increase to 40% and 30% and so we see that the identified set is larger. As we can, the model with non-stationarity still contains information about the parameters of interest. We also simulated cases with at least 50% censoring that resulted in a model with no information about β as our results above suggest.

6.3 Identified Set in Dynamic Models

Here, we first simulate the following dynamic model in which a *lagged observed variable* is on the right hand side:

$$y_{it}^* = \gamma_0 v_{it-1} + x_{it}'\beta_0 + \alpha_i + \epsilon_{it}$$

Here, we assume that the initial period is observed, is $\mathcal{N}(0, 1)$ and is independent of all variables in the model. In addition, we simulate the fixed effects and the errors as above. On top of Figure 4, we have the model censored at -1 which resulted in almost 30% censoring in each period. For the random independent censoring case, we use random normal censoring with mean -1, and for the endogenous censoring we have $c_{it} = \mathcal{N}(-1, 1) + .2\epsilon_{it}$. In addition, the covariate dependent model adds the sum of the covariates across time periods to c_{it}. As we can see, the presence of lagged y_{it} does not result in a complete lack of identification for the above model.

Next, we turn to the dynamic model with lagged sector specific variables as regressors which is provided in Figure 5. There, we plot the identified set for (β, γ) in the following model:

$$y_{it}^* = \alpha_i + \gamma_0 d_{it} + x_{it}' \beta_0 + \epsilon_{it}$$

where again, $d_{it} = 1[y_{it}^* \geq c_{it}]$, an observed binary sector indicator variable. The model is simulated with the same values as the previous models. As we can see from the plots in Figure 4, the sizes of the identified set seems similar and more importantly, it is clear that a stationary dynamic model does not generally identify the parameter of interest in this design, but do contain information.

6.4 Monte Carlo Evidence: Confidence Regions

In this section, we use our inference results to construct confidence regions for parameters in the stationary model that we used above:

$$y_t^* = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \quad t = 1, 2$$

where $\beta_1 = \beta_0 = 1$ with varying sample sizes and under different assumptions on the censoring mechanism. In particular, we first calculate $\hat{P}\{\mathbf{x}_i = \mathbf{x}\}$, then, we calculate for a given parameter vector $\mathbf{b}$, $\hat{D}_1(\tau, \mathbf{x}; \mathbf{b})$ and $\hat{D}_2(\tau, \mathbf{x}; \mathbf{b})$ and evaluate those on a sequence of random τ grids from -4 to 4 of size 10. We then calculate $\hat{\delta}_{i1}(\tau, \mathbf{x}, \mathbf{b})$, $\hat{\delta}_{i2}(\tau, \mathbf{x}, \mathbf{b})$ and $T_n(\beta)$. To get the critical values, we simulate the gaussian process G^{NS} in a simple way as follows. We simulate $\{u_{is}, i = 1, \ldots, n\}$ from a standard normal distribution. Then, for $j = 1, 2$:

$$\hat{G}_{s,j}(\tau, \mathbf{x}; \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\hat{\delta}_{ij}(\tau, \mathbf{x}; \beta) - \hat{\Delta}_1(\tau, \mathbf{x}; \beta)) u_{is}$$

and then we construct the $(1-\alpha).100\% CS$ for β in the following way:

$$CS_{n,1-\alpha} = \{\beta \in B : T_n(\beta) \leq c_{n,1-\alpha}^S(\beta)\}$$

where $c_{n,1-\alpha}^S(\beta)$ is the $(1-\alpha).100\%$ quantile of the empirical distribution of

$$\sup_{(\tau, \mathbf{x}) \in R \times \chi} \max\{\hat{G}_{s,1}(\tau, \mathbf{x}; \beta), \hat{G}_{s,2}(\tau, \mathbf{x}; \beta)\}$$

The way we do that is we construct a grid for (β_1, β_2) and then performing the test pointwise. We report in the tables, projections of the confidence region onto the two axes. So, the coverage of these "rectangular" confidence regions is at least $(1-\alpha).100\%$ where here we set that equal to 95%. The various *Models* we report are simulated similar to the previous section, so the *Random Independent Censoring* model is one where c is $\mathcal{N}(0, .25)$, while the *Random Covariate Dependent Censoring* is one where c_1 is $\mathcal{N}(0,1) + (x_{21} - x_{11})$. As we can see from Table 1, our estimation approach seem to perform well[10] in small samples. Here, again, the CI were constructed by inverting the test statistic T_n and for example in the fixed censoring case where we know the model is point identified, the (conservative) confidence regions are well behaved but tend to be slightly non-symmetric with respect to the truth of $(1, 1)$.

7 Conclusion

This paper considers identification and inference in a class of censored models in panel data settings. Our main contribution is to provide the tightest sets on the parameter of interest that we can learn from data at hand under two sets of assumptions. Throughout, we allow the censoring to be completely general with no restrictions on the relationship between the censoring variable and the other variables in the model. In the specific setting resulting in a randomly censored regression model our results nest existing work on censoring in both panel and cross section settings, such as Honoré (1992), Honoré, Khan, and Powell (2002), and Honoré and Powell (1994). The paper also contains novel results on identification for *dynamic models* where various kinds of "lagged" behavior is allowed such as having a lagged indicator, a lagged observed outcome, and a lagged latent outcome. In addition, we provide characterizations of the identified set in a model with factor loads. The area of panel data Roy models with dynamics is not well understood in the literature as conditions for point identification under reasonable assumptions are not available. Hence, our results provide a step in that direction in that we construct the identified set for such dynamic models under weak assumptions.

In addition, our characterization of the identified sets are constructive in that they can

[10] The simulations were all conducted using Matlab on a generic office computer and the time each took ranged from 1.5 hours to 8 hours.

Table 1: 95% Confidence Regions for Stationary Model $n = 180, 360, 600$, $T = 2$.

Model	Marginal 95% CI	
	β_1	β_2
Fixed Censoring		
$N = 180$	$[.4, 2.4]$	$[-.2, 2.6]$
$N = 360$	$[.7, 1.7]$	$[.5, 1.9]$
$N = 600$	$[.85, 1.45]$	$[.75, 1.76]$
Random Independent Censoring		
$N = 180$	$[.35, 2.3]$	$[-.21, 2.9]$
$N = 360$	$[.72, 1.66]$	$[.54, 1.92]$
$N = 600$	$[.91, 1.15]$	$[.85, 1.26]$
Random Covariate Dependent Censoring		
$N = 180$	$[.35, 2.3]$	$[-.21, 2.9]$
$N = 360$	$[.72, 1.72]$	$[.5, 1.9]$
$N = 600$	$[.83, 1.29]$	$[.81, 1.32]$
Random Endogenous Censoring		
$N = 180$	$[.35, 2.3]$	$[-.21, 2.9]$
$N = 360$	$[.7, 1.7]$	$[.5, 1.9]$
$N = 600$	$[.761.25]$	$[.75, 1.46]$

be estimated from the sample. We then provide a practical approach to constructing confidence regions that control size and can be used to get correct confidence regions. The proposed inference method is based on conditional moment inequalities that is *adaptive* to point identification conditions in the sense that the objective function is minimized at the identified set (which can be a singleton in the point identified case), depending on the features of the data generating process. We also provide guidance on how one might construct confidence regions for the identified set based on recent contributions to the theory of stochastic dominance tests (for example, Jun, Lee, and Shin (2011), Linton, Maasoumi, and Whang (2005) and Linton, Song, and Whang (2010)).

References

Andrews, D., and X. Shi (2013): "Inference based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Arellano, M. (2003): *Panel data econometrics*. Oxford University Press, USA.

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 58(2), 277.

Arellano, M., and S. Bonhomme (2009): "Robust priors in nonlinear panel data models," *Econometrica*, 77(2), 489–536.

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

Baltagi, B. (1995): *Econometric Analysis of Panel Data*. New York, NY: Wiley.

Beresteanu, A., F. Molinari, and I. Molchanov (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79, 1785–1821.

Bester, A., and C. Hansen (2009): "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business and Economic Statistics*, 27(2), 235–250.

Bonhomme, S. (2012): "Functional Differencing," *Econometrica*, 80, 1337–1385.

Chay, K., and B. Honoré (1998): "Estimation of Semiparametric Censored Regression Models: An Application to Changes in Black-White Earnings Inequality During the 1960's.," *Journal of Human Resources*, 33, 4–38.

Chen, S., and S. Khan (2008): "Semiparametric Estimation of Non-stationary Panel Data Censored Regression Models with Time Varying Factor Loads," *Econometric Theory*, 24, 1149–1173.

Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2010): "Identification and Estimation of Marginal Effects in Nonlinear Panel Models," manuscript.

Chernozhukov, V., S. Lee, and A. Rosen (2013): "Interesection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Evdokimov, K. (2010): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," Working Paper.

Graham, B., and J. Powell (2012): "Identification and Estimation of 'Irregular' Correlated Random Coefficient Models," *Econometrica*, 80, 2105–2152.

Hoderlein, S., and H. White (2012): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *Journal of Econometrics*, 168, 300–314.

Honoré, B. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60(3), 533–65.

Honoré, B. (1993): "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *Journal of Econometrics*, 59, 35–61.

Honoré, B., and L. Hu (2004): "Estimation of Cross sectional and Panel Data Censored Regres- sion Models with Endogeneity," *Journal of Econometrics*, 122, 293–316.

Honoré, B., S. Khan, and J. Powell (2002): "Quantile Regression Under Random Censoring," *Journal of Econometrics*, 64, 241–278.

Honoré, B., and A. Lleras-Muney (2006): "Bounds in Competing Risks Models and the War on Cancer," *Econometrica*, (6), 1675–1698.

Honoré, B., and J. Powell (1994): "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, 64, 241–278.

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

Hsiao, C. (1986): *Analysis of Panel Data*. Cambridge University Press.

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

Jun, S., Y. Lee, and Y. Shin (2011): "Testing for Distributional Treatment Effects: A Set Identification Approach," Working Paper.

Khan, S., and E. Tamer (2007): "Partial Rank Estimation of Transformation Models with General forms of Censoring," *Journal of Econometrics*, 136, 251–280.

——— (2009): "Inference on Endogenously Censored Regressions Models Using Conditional Moment Inequalities," *Journal of Econometrics*, 152, 104–119.

——— (2010): "Irregular Identification, Support Conditions and Inverse Weight Estimation," *Econometrica*, 78, 1783–2103.

Kim, K. (2007): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," University of Minnesota Working Paper.

Linton, O., E. Maasoumi, and Y. Whang (2005): "Consistent Testing for Stochastic Dominance under General Sampling Schemes," *Review of Economic Studies*, 72(3), 735–765.

Linton, O., K. Song, and Y. Whang (2010): "An Improved Bootstrap Test for Stochastic Dominance," *Journal of Econometrics*, 154(2), 186–202.

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.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519–547.

Moon, R., and P. Phillips (2000): "Nonstationary Panel Data Analysis: An Overview of Some Recent Developments," *Econometric Reviews*, 19, 263–286.

Ponomareva, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," Working Paper.

Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, pp. 303–325.

Rosen, A. (2012): "Set Identification via Quantile Restrictions in Short Panels," *Journal of Econometrics*, 166, 127–137.

A Appendix

A.1 Point Identification

A.1.1 Point Identification in a Stationary Model

In this section we establish sufficient conditions for point identification in a stationary model: we show that under certain assumptions one can come up with a set of moment conditions that point identify the parameter of interest. Those moment conditions are similar to the moment conditions used maximum score estimator in Manski (1985) or in partial rank estimator in Khan and Tamer (2007).

We assume that the following conditions hold.

Assumption 12 (Large Support). *Conditional on all other components (denoted by subscript $-k$), the distribution of k^{th} component of vector Δx_i is absolutely continuous on $\mathbb{R}$ with respect to Lebesgue measure, $supp(\Delta x_{i,k}|\Delta x_{i,-k}) = \mathbb{R}$, and $\beta_k \neq 0$. Also, the support of Δx_i is not contained in any proper linear subspace of $\mathbb{R}^k$.*

Assumption 13 (Bounded Censoring). *There exist (random) variables τ_{i1} and τ_{i2} with a known distribution such that the τ_{it} is independent of $\mathbf{x}_i, \epsilon_{it}, \alpha_i$, and c_{it}, τ_{i1} and τ_{i2} are independent, and the set $\Xi = \{\mathbf{x} \in \mathcal{X} : \quad P(c_{it} \leq \tau_{it}|\mathbf{x}_i = \mathbf{x}) = 1\}$ is non-empty. Also, $\inf_{\mathbf{x}\in\Xi} P\{d_{it} = 1|\mathbf{x}_i = \mathbf{x}\} > 0$.*

Assumption 12 is a standard identifying condition used in in settings where maximum-score type settings. Assumption 13 ensures that we can come up with some *exogenous* censoring procedure that dominates the endogenous censoring at least on some subset of the support of $\mathbf{x}_i$. For example, Assumption 13 holds if censoring variable c_{it} is endogenous, but bounded above by some constant M. Another example would be the case when censoring is exogenous: then we can choose $\tau_{it} = c_{it}$. Second part of Assumption 13 makes sure that there is not too much censoring.

Theorem A.1 (Point Identification in a Stationary Model). *Assume that Assumptions 1, 12 and 13 hold. Let $F_\eta(\cdot|\mathbf{x}_i)$ denote the conditional distribution of $\alpha_i + \epsilon_{it}$ conditional on $\mathbf{x}_i$ and assume that $F_\eta(\cdot|\mathbf{x}_i = \mathbf{x})$ is strictly increasing for each $\mathbf{x}$ in the support of $\mathbf{x}_i$. Then $B_I = \{\beta\}$ and so β is point identified.*

Theorem A.1 implies that one can consistently estimate β using a maximum rank correlation type estimator.

A.1.2 Point Identification in a Non-Stationary Model

In this section we establish sufficient conditions for point identification in a non-stationary model. As we noted previously, if no censoring occurs for a subset of the support of $\mathbf{x}_i$ such that the corresponding subset of the support of Δx_i is not contained in any proper linear subspace of $\mathbb{R}^k$, then $B_I = \{\beta\}$. However, it is possible to point identify β or some components of it without requiring y^*_{it} being fully observed in both period for a subset of the support of $\mathbf{x}_i$.

We start by defining

$$p(\mathbf{x}_i) = P\{y^*_{i1} > c_{i1}, y^*_{i2} > c_{i2} | \mathbf{x}_i\}$$

Quantity $p(\mathbf{x}_i)$ represents the fraction of population that is uncensored in both periods for a set of covariates $\mathbf{x}_i$. Then, given the definition of the upper bound $UB(\tau, \mathbf{x}_i, b)$ in Theorem 3.1, we have

$$UB(\tau, \mathbf{x}_j, b) \leq P\{\Delta\epsilon_j \leq \tau + \Delta x'_j(b - \beta)|\mathbf{x}_j\} + 1 - p(\mathbf{x}_j)$$

Similarly, given the definition of the lower bound in Theorem 3.1, we get the following inequality for the lower bound:

$$LB(\tau, \mathbf{x}_i, b) \geq P\{\Delta\epsilon_i \leq \tau + \Delta x'_i(b - \beta)|\mathbf{x}_i\} - 1 + p(\mathbf{x}_i)$$

Therefore, for all $b \in B_I$ it must hold that

$$F_{\Delta\epsilon}(\tau + \Delta x'_i(b - \beta)|\mathbf{x}_i) - F_{\Delta\epsilon}(\tau + \Delta x'_j(b - \beta)|\mathbf{x}_j) \leq 2 - p(\mathbf{x}_i) - p(\mathbf{x}_j) \quad \text{(A.1)}$$

for all τ, $\mathbf{x}_i$ and $\mathbf{x}_j$, where $F_{\Delta\epsilon}(\cdot|\mathbf{x}_i)$ denotes the conditional distribution of $\Delta\epsilon_i$ given $\mathbf{x}_i$. This motivates the following *sufficient condition* for point identification of β.

Assumption 14 ("Not Too Much Censoring").

(i) There exists $0 < q < 1$ *such that for all* $\mathbf{x}_i$, $\mathbf{x}_j$ *it holds that* $2 - p(\mathbf{x}_i) - p(\mathbf{x}_j) < q$.

(ii) For any $\mathbf{x}_{i,-k}$, $\sup_{\mathbf{x}_{i,k}\in\mathbb{R}} p(\mathbf{x}_{i,k}, \mathbf{x}_{i,-k}) = 1$.

For example, if for each $\mathbf{x}_i$ at least 51% or more of observations are uncensored in both periods, then condition (i) with any q between .98 and 1. Further, we want to note that the first part of Assumption 14 is partially testable (in that we can test the null hypothesis that $2 - p(\mathbf{x}_i) - p(\mathbf{x}_j) < q$ if we fix some small value q).

The following theorem uses identification at infinity argument to point identify either β or its k^{th} component. Note that Assumption 14 is by no means necessary for point identification.

Theorem A.2 (Point Identification in a Non-Stationary Model). *Let Assumptions 3, 12 and 14(i) hold, and suppose that* $b \in B$ *is such that* $b_k \neq \beta_k$*. Then*

1. β *is identified relative to* b.
2. *Additionally, if assumption Assumption 14((ii) holds, then* β *is point identified, so that* $B_I = \{\beta\}$.

The point identification result above relies on variation at infinity to shrink the set B_I to a single point. Notice that although it requires large supports, this type of point identification is robust in that if in fact the regressors do not have large support, the identified set is non-trivial as was shown previously. For more on robust point identification, See Khan and Tamer (2010).

A.2 Stationary Model with Lagged Latent Outcome

In this section, we characterize the sharp set in the lagged outcome model. The characterization here is tedious since the *latent* y's (the y^*'s) are included on the rhs which makes the model harder. In essence, the reason why is because here y_1^* which is generally censored not only appears in the lhs in time period 1, but it also appears on the rhs in $T = 2$. So, deriving the identified set must take account of that.

If we observe the latent outcome in the first period, y_{1i}^*, then

$$\alpha_i + \epsilon_{i1} = y_{i1}^* - \gamma y_{i0}^* - x_{i1}'\beta$$

The upper and lower bounds on $\alpha_i + \epsilon_{i12}$ are, as before,

$$y_{i2}^L - \gamma y_{i1}^* - x_{i2}'\beta \le \alpha_i + \epsilon_{i2} \le y_{i2}^U - \gamma y_{i1}^* - x_{i2}'\beta$$

If conditional on $\mathbf{x}_i$ and α_i the distribution of ϵ_{it} is stationary, the following inequalities must hold[11]:

$$P\{y_{i2}^U - \gamma y_{i1}^* - x_{i2}'\beta|\mathbf{x}_i\} \le P\{y_{i1}^* - \gamma y_{i0}^* - x_{i1}'\beta|\mathbf{x}_i\} \le P\{y_{i2}^L - \gamma y_{i1}^* - x_{i2}'\beta|\mathbf{x}_i\}$$

Therefore, the sharp identified set here can be characterized as follows: let $\tilde{y}_i$ be a random variable with the (continuous conditional on $\mathbf{x}_i$) distribution such that

$$\tilde{y}_i = \begin{cases} y_{i1} & \text{if } y_{i1}^* > c_{i1} \\ u_i \le c_{i1} & \text{if } y_{i1}^* \le c_{i1} \end{cases}$$

where the u_i is a continuous random variable with distribution that can potentially depend on $\mathbf{x}_i$ and has to obey the above support conditions. Let $(g, b) \in \Theta = \Gamma \times B$ be such that the inequalities

$$P\{y_{i2}^U - g\tilde{y}_i - x_{i2}'b|\mathbf{x}_i\} \le P\{\tilde{y}_i - g y_{i0}^* - x_{i1}'b|\mathbf{x}_i\} \le P\{[y_{i2}^L - g\tilde{y}_i - x_{i2}'b|\mathbf{x}_i\} \tag{A.2}$$

hold for all $\mathbf{x}_i$ in the support of $\mathbf{x}_i$. Then (g, b) is observationally equivalent to (γ, β).

Unfortunately, the probabilities in (A.2) depend on the choice of the distribution for $\tilde{y}_i$, and hence those inequalities cannot be thought of as a set of conditional moment inequalities that can be calculated from the data once parameters g and b are given. An approach to building the sharp set in this case would be collect (g, b)'s that correspond *to all* (conditional) distributions u_i that satisfy the above support conditions.

A.3 Proofs

A.3.1 Proof of Theorem 2.1

Suppose that $b \in B_I$. We will construct $\tilde{y}_{it}^*$ and $\tilde{c}_{it}$ such that *(i)* $\tilde{y}_{it} = \max\{\tilde{y}_{it}^*, \tilde{c}_{it}\}$ has the same distribution conditional on $\mathbf{x}_i$ as y_{it} for $t = 1, 2$ and *(ii)* $\tilde{y}_{it}^* = x_{it}'b + \tilde{\alpha}_i + \tilde{\epsilon}_{it}$, where $\tilde{\alpha}_i + \tilde{\epsilon}_{i1}$ and $\tilde{\alpha}_i + \tilde{\epsilon}_{i2}$ are identically distributed conditional on $\mathbf{x}_i$. For the ease of presentation, we define $\eta_{it} \equiv \alpha_i + \epsilon_{it}$ and $\tilde{\eta}_{it} \equiv \tilde{\alpha}_i + \tilde{\epsilon}_{it}$.

[11] This is for $T = 2$. A similar procedure works for $T > 2$.

Note that

$$P\{y_{it}^L - x_{it}'b \leq \tau|\mathbf{x}_i\} = P\{\eta_{it} \leq \tau + x_{it}'(b-\beta), y_{it}^* > c_{it}|\mathbf{x}_i\} + P\{y_{it}^* \leq c_{it}|\mathbf{x}_i\}$$

and

$$P\{y_{it}^U - x_{it}'b \leq \tau|\mathbf{x}_i\} = P\{\eta_{it} \leq \tau + x_{it}'(b-\beta), y_{it}^* > c_{it}|\mathbf{x}_i\} + P\{c_{it} - x_{it}'b \leq \tau, y_{it}^* \leq c_{it}|\mathbf{x}_i\}$$

Let $\tilde{c}_{it} = c_{it}$ and define $\tilde{\eta}_{it}$ as follows:

- If $y_{it}^* \geq c_{it}$: $\tilde{\eta}_{it} = \eta_{it} + x_{it}'(\beta - b)$.
- If $y_{it}^* < c_{it}$: $\tilde{\eta}_{it} = u_{it} < c_{it} - x_{it}'b$, where u_{it} is a random variable that can depend on x_{it}, c_{it}, and η_{it}.

In this case, $\tilde{y}_{it} = y_{it}$ and $\tilde{d}_{it} = d_{it}$ for $t = 1,2$, where $\tilde{d}_{it} = 1\{\tilde{y}_{it}^* \geq \tilde{c}_{it}\}$. We want $P\{\tilde{\eta}_{i1} \leq \tau|\mathbf{x}_i\} = P\{\tilde{\eta}_{i2} \leq \tau|\mathbf{x}_i\}$. For each $t = 1,2$, the sharp upper bound on $P\{\tilde{\eta}_{it} \leq \tau|\mathbf{x}_i\}$ is $P\{\eta_{it} \leq \tau + x_{it}'(b-\beta), y_{it}^* \geq c_{it}|\mathbf{x}_i\} + P\{y_{it}^* < c_{it}|\mathbf{x}_i\} = P\{y_{it}^L - x_{it}'b \leq \tau|\mathbf{x}_i\}$, while the sharp lower bound (over all possible distributions of u_{it} such that $u_{it} < c_{it} - x_{it}'b$) is $P\{\eta_{it} \leq \tau + x_{it}'(b-\beta), y_{it}^* \geq c_{it}|\mathbf{x}_i\} + P\{c_{it} - x_{it}'b \leq \tau, y_{it}^* < c_{it}|\mathbf{x}_i\} = P\{y_{it}^U - x_{it}'b \leq \tau|\mathbf{x}_i\}$. Any (continuous) distribution between these upper and lower bounds can be generated by some distribution of u_{it}. Finally, since b satisfies conditional inequalities (2.2), then we can find u_{i1} and u_{i2} distributed in such a way that $P\{\tilde{\eta}_{i1} \leq \tau|\mathbf{x}_i\} = P\{\tilde{\eta}_{i2} \leq \tau|\mathbf{x}_i\}$ (i.e., those composite error terms satisfy Assumption 1). Therefore, b is observationally equivalent to β. □

A.3.2 Proof of Theorem 3.1

We can re-write lower bound as $LB(\tau, \mathbf{x}_i, b) = P\{y_{i2}^U - y_{i1}^L - \Delta x_i'b \leq \tau|\mathbf{x}_i\} = P\{y_{i2}^U - y_{i1}^L - \Delta x_i'b \leq \tau, y_{i2}^* \geq c_{i2}, y_{i1}^* \geq c_{i1}\} + P\{y_{i2}^U - y_{i1}^L - \Delta x_i'b \leq \tau, y_{i2}^* \geq c_{i2}, y_{i1}^* < c_{i1}|\mathbf{x}_i\} + P\{y_{i2}^U - y_{i1}^L - \Delta x_i'b \leq \tau, y_{i2}^* < c_{i2}, y_{i1}^* \geq c_{i1}|\mathbf{x}_i\} + P\{y_{i2}^U - y_{i1}^L - \Delta x_i'b \leq \tau, y_{i2}^* < c_{i2}, y_{i1}^* < c_{i1}|\mathbf{x}_i\} = P\{\Delta\epsilon_i + \Delta x_i'\beta \leq \tau + \Delta x_i'b, y_{i2}^* \geq c_{i2}, y_{i1}^* \geq c_{i1}|\mathbf{x}_i\} + 0 + P\{c_{i2} - y_{i1}^* \leq \tau + \Delta x_i'b, y_{i2}^* < c_{i2}, y_{i1}^* \geq c_{i1}|\mathbf{x}_i\} + 0$. So that

$$\begin{aligned} LB(\tau, \mathbf{x}_i, b) = \quad & P\{\Delta\epsilon_i + \Delta x_i'\beta \leq \tau + \Delta x_i'b, y_{i2}^* \geq c_{i2}, y_{i1}^* \geq c_{i1}|\mathbf{x}_i\} \\ & + P\{c_{i2} - y_{i1}^* \leq \tau + \Delta x_i'b, y_{i2}^* < c_{i2}, y_{i1}^* \geq c_{i1}|\mathbf{x}_i\} \end{aligned} \tag{A.3}$$

Similarly, we can re-write upper bound as $UB(\tau, \mathbf{x}_j, b) = P\{y^L_{j2} - y^U_{j1} - \Delta x'_j b \le \tau | \mathbf{x}_j\} = P\{y^L_{j2} - y^U_{j1} - \Delta x'_j b \le \tau, y^*_{j2} \ge c_{j2}, y^*_{j1} \ge c_{j1} | \mathbf{x}_j\} + P\{y^L_{j2} - y^U_{j1} - \Delta x'_j b \le \tau, y^*_{j2} \ge c_{j2}, y^*_{j1} < c_{j1} | \mathbf{x}_j\} + P\{y^L_{j2} - y^U_{j1} - \Delta x'_j b \le \tau, y^*_{j2} < c_{j2}, y^*_{j1} \ge c_{j1} | \mathbf{x}_j\} + P\{y^L_{j2} - y^U_{j1} - \Delta x'_j b \le \tau, y^*_{j2} < c_{j2}, y^*_{j1} < c_{j1} | \mathbf{x}_j\} = P\{\Delta\epsilon_j + \Delta x'_j \beta \le \tau + \Delta x'_j b, y^*_{j2} \ge c_{j2}, y^*_{j1} \ge c_{j1} | \mathbf{x}_j\} + P\{y^*_{j2} - c_{j1} \le \tau + \Delta x'_j b, y^*_{j1} < c_{j1}, y^*_{j2} \ge c_{j2} | \mathbf{x}_j\} + P\{y^*_{j1} \ge c_{j1}, y^*_{j2} < c_{j2} | \mathbf{x}_j\} + P\{y^*_{j1} < c_{j1}, y^*_{j2} < c_{j2} | \mathbf{x}_j\}$. So that

$$
\begin{aligned}
UB(\tau, \mathbf{x}_j, b) = \quad & P\{\Delta\epsilon_j + \Delta x'_j \beta \le \tau + \Delta x'_j b, y^*_{j2} \ge c_{j2}, y^*_{j1} \ge c_{j1} | \mathbf{x}_j\} \\
& + P\{y^*_{j2} < c_{j2} | \mathbf{x}_j\} + P\{y^*_{j2} - c_{j1} \le \tau + \Delta x'_j b, y^*_{j2} \ge c_{j2}, y^*_{j1} < c_{j1} | \mathbf{x}_j\}
\end{aligned} \tag{A.4}
$$

Suppose that $b \in B_I$, that is

$LB(\tau, \mathbf{x}_i, b) \le UB(\tau, \mathbf{x}_j, b)$ for all τ, $\mathbf{x}_i$, $\mathbf{x}_j$.

Now let $\tilde{c}_{i1} = c_{i1}$, $\tilde{c}_{i2} = c_{i2}$ and define $\Delta\tilde{\epsilon}_i$ and $\tilde{\alpha}_i$ as follows:

- If $y^*_{i2} \ge c_{i2}$, $y^*_{i1} \ge c_{i1}$, then $\tilde{\alpha}_i = \alpha_i + x'_{i1}\beta - x'_{i1} b$, and $\Delta\tilde{\epsilon}_i = \Delta\epsilon_i + \Delta x'_i \beta - \Delta x'_i b$.
- If $y^*_{i2} \ge c_{i2}$, $y^*_{i1} < c_{i1}$, then $\tilde{\alpha}_i = y^*_{i2} - \Delta\tilde{\epsilon}_i - x'_{i2} b$, and $\Delta\tilde{\epsilon}_i = \gamma_i(\Delta\epsilon_i + \Delta x'_i \beta) + (1 - \gamma_i)(y^*_{i2} - c_{i1}) - \Delta x'_i b + u_{i1}$, where $0 \le \gamma_i \le 1$ and $u_{i1} > 0$.
- If $y^*_{i2} < c_{i2}$, $y^*_{i1} \ge c_{i1}$, then $\tilde{\alpha}_i = \alpha_i + x'_{i1}\beta - x'_{i1} b$, and $\Delta\tilde{\epsilon}_i = \lambda_i(\Delta\epsilon_i + \Delta x'_i \beta) + (1 - \lambda_i)(c_{21} - y^*_{i1}) - \Delta x'_i b - u_{i2}$, where $0 \le \lambda_i \le 1$ and $u_{i2} > 0$.
- If $y^*_{i2} < c_{i2}$, $y^*_{i1} < c_{i1}$, then $\Delta\tilde{\epsilon}_i = \Delta\epsilon_i + \Delta x'_i \beta - \Delta x'_i b - u_{i3}$ and $\tilde{\alpha}_i = \min\{c_{i1} - x'_{i1} b, c_{i2} - \Delta\epsilon_i - \Delta x'_i \beta + \Delta x'_i b + u_{i3}\} - u_{i4}$, where $-\infty < u_{i3} < +\infty$ and $u_{i4} > 0$.

Here $u_{i1}, u_{i2}, u_{i3}, u_{i4}$, λ_i, and γ_i are random variables that may depend on $\mathbf{x}_i$, $\Delta\epsilon_i$, α_i etc. Let $\tilde{y}_{i1} = \max\{x'_{i1} b + \tilde{\alpha}_i + \epsilon_{i1}, \tilde{c}_{i1}\}$ and $\tilde{y}_{i2} = \max\{x'_{i2} b + \tilde{\alpha}_i + \Delta\tilde{\epsilon}_i, \tilde{c}_{i2}\}$. Then $(\tilde{y}_{i1}, \tilde{y}_{i2}) = (y_{i1}, y_{i2})$.

Now, $P\{\Delta\tilde{\epsilon}_i \le \tau | \mathbf{x}_i\} = P\{\Delta\epsilon_i + \Delta x'_i \beta \le \tau + \Delta x'_i b, y^*_{i2} \ge c_{i2}, y^*_{i1} \ge c_{i1} | \mathbf{x}_i\} + P\{\gamma_i(\Delta\epsilon_i + \Delta x'_i \beta) + (1 - \gamma_i)(y^*_{i2} - c_{i1}) \le \tau + \Delta x'_i b - u_{i1}, y^*_{i2} \ge c_{i2}, y^*_{i1} < c_{i1} | \mathbf{x}_i\} + P\{\lambda_i(\Delta\epsilon_i + \Delta x'_i \beta) + (1 - \lambda_i)(c_{i1} - y^*_{i1}) \le \tau + \Delta x'_i b + u_{i2}, y^*_{i2} < c_{i2}, y^*_{i1} \ge c_{i1} | \mathbf{x}_i\} + P\{\Delta\epsilon_i + \Delta x'_i \beta \le \tau + \Delta x'_i b + u_{i3}, y^*_{i2} < c_{i2}, y^*_{i1} < c_{i1} | \mathbf{x}_i\}$.

Then lower (sharp) bound on $P\{\Delta\tilde{\epsilon}_i \le \tau | \mathbf{x}_i\}$ over all possible distributions of u_{i1}, u_{i2}, u_{i3}, u_{i4}, λ_i, and γ_i is equal to $LB(\tau, \mathbf{x}_i, b)$, and upper (sharp) bound on $P\{\Delta\tilde{\epsilon}_j \le \tau | \mathbf{x}_j\}$ is equal to $UB(\tau, \mathbf{x}_j, b)$. Therefore, it is possible to find such a distribution of $u_{i1}, u_{i2}, u_{i3}, u_{i4}$, λ_i, and γ_i (conditional on $\mathbf{x}_i$ etc) so that for every τ, $\mathbf{x}_i$, and $\mathbf{x}_j$ we have $P\{\Delta\tilde{\epsilon}_i \leq \tau|\mathbf{x}_i\} = P\{\Delta\tilde{\epsilon}_i \leq \tau|\mathbf{x}_j\} = F(\tau)$ for some $F(\tau)$ such that $LB(\tau, \mathbf{x}_i, b) \leq F(\tau) \leq UB(\tau, \mathbf{x}_j, b)$, and this distribution is independent of $\mathbf{x}_i$. That is, we constructed $\tilde{y}_{it}$ and $\tilde{c}_{it}$ such that: (a) vector $(\tilde{y}_{i1}, \tilde{y}_{i2}, \tilde{d}_{i1}, \tilde{d}_{i2})$ is distributed as $(y_{i1}, y_{i2}, d_{i1}, d_{i2})$; and (b) Assumption 3 (or 4) is satisfied for $\tilde{y}^*_{i1} = x_{i1}b + \tilde{\alpha}_i + \epsilon_{i1}$ and $\tilde{y}^*_{i2} = x_{i2}b + \tilde{\alpha}_i + \Delta\tilde{\epsilon}_i$. □

A.3.3 Proof of Theorem 3.2.

Proof: Let $w_i^L = y_{i2}^L - y_{i1}^U$ and $w_i^U = y_{i2}^U - y_{i1}^L$. Then the sharp identified set can be written as $B_I = \{b : \text{for every } \tau, \mathbf{x}_i, \mathbf{x}_j \; P\{w_i^U - \Delta x_i' b \leq \tau|\mathbf{x}_i\} \leq P\{w_j^L - \Delta x_j' b \leq \tau|\mathbf{x}_j\}\}$. Note that for every τ_1, τ_2, $P\{w_j^L - \Delta x_j' b \leq \tau_1|\mathbf{x}_j\} \geq p_2^c(\mathbf{x}_j)$ and $P\{w_i^U - \Delta x_i' b \leq \tau_2|\mathbf{x}_i\} \leq 1 - p_2^c(\mathbf{x}_i)$. Therefore, if $1 - p_2^c(\mathbf{x}_i) \leq p_1^c(\mathbf{x}_j)$ for all $\mathbf{x}_i$ and $\mathbf{x}_j$, then we have $P\{w_i^U - \Delta x_i' b \leq \tau|\mathbf{x}_i\} \leq 1 - p_1^c(\mathbf{x}_i) \leq p_1^c(\mathbf{x}_j) \leq P\{w_j^L - \Delta x_j' b \leq \tau|\mathbf{x}_j\}$ for every $b \in B$, so the bounds are trivial. □

A.3.4 Proof of Theorem 3.3

Note first that for all b, $Q(b) \geq 0$, so that $B_Q = \arg\min_b Q(b)$. Next, let $b \in B_I$ and recall that B_I is defined by the following set of inequalities:

$$P\{w_i^U - \Delta x_i' b \leq \tau|\mathbf{x}_i\} \leq F(\tau) \leq P\{w_j^L - \Delta x_j' b \leq \tau|\mathbf{x}_j\}\} \tag{A.5}$$

for some cumulative distribution function F. Inequalities (A.5) imply that if for some constants τ_1 and τ_2 we have $\tau_2 - \Delta x_j' b \geq \tau_1 - \Delta x_i' b$, then $P\{w_i^U \leq \tau_1|\mathbf{x}_i\} \leq P\{w_j^L \leq \tau_2|\mathbf{x}_j\}$. Therefore, if $b \in B_I$, then $Q(b) = 0$, so that $B_I \subseteq B_Q$.

Now suppose that there exists $b \in B_Q$ such that $b \notin B_I$. That is, for this b there exist $\tilde{\tau}$, $\tilde{x}_i$ and $\tilde{x}_j$ such that

$$P\{w_i^U \leq \tilde{\tau} + \Delta\tilde{x}_i b|\tilde{x}_i\} > P\{w_j^L \leq \tilde{\tau} + \Delta\tilde{x}_j b|\tilde{x}_j\} \tag{A.6}$$

Let $\tilde{\tau}_{2j} = \tilde{\tau} + \Delta\tilde{x}_j b$ and $\tilde{\tau}_{1i} = \tilde{\tau} + \Delta\tilde{x}_i b$. Then $\tilde{\tau}_{2j} - \Delta\tilde{x}_j b = \tilde{\tau}_{1i} - \Delta\tilde{x}_i b = \tilde{\tau}$ and $P\{\Delta w_{ui} \leq \tilde{\tau}_{1i}|\tilde{x}_i\} > P\{\Delta w_{lj} \leq \tilde{\tau}_{2j}|\tilde{x}_j\}$. By continuity of τ and strict inequality in (3.1), there exist the set U of positive probability measure such that for all $(\tau_{1i}, \tau_{2i}, \mathbf{x}_i, \mathbf{x}_j) \in U$ we have:

1. $\tau_{2j} - \Delta x_j' b \geq \tau_{1i} - \Delta x_i' b$,

2. $P\{\Delta w_{ui} \leq \tau_{1i} | \mathbf{x}_i\} > P\{\Delta w_{lj} \leq \tau_{2j} | \mathbf{x}_j\}$,

so that $Q(b) > 0$, which implies that if $b \notin B_I$, then $Q(b) > 0$. Therefore, $B_I = B_Q$. □

A.3.5 Proof of Theorem 3.4

The proof uses similar arguments as the ones in Theorem 2.1 or Theorem 3.1 and therefore is omitted.

A.3.6 Proof of Theorem 4.1

First, of (γ, β) are the true parameter values, then by construction $(\gamma, \beta) \in \Theta_{I,1}$. Let $\eta_{it} = \alpha_i + \epsilon_{it}$, for $t = 1, 2$. For any $(g, b) \in \Theta_{I,1}$, let $\tilde{y}^*_{it} = g y_{i,t-1} + x'_{it} b + \tilde{\eta}_{it}$, $\tilde{c}_{it} = c_{it}$. For $\tilde{\eta}_{it}$ we have:

- If $y^*_{it} \geq c_{it}$: $\tilde{\eta}_{it} = \eta_{it} + (\gamma - g) y_{i,t-1} + x'_{it}(\beta - b)$.
- If $y^*_{it} < c_{it}$: $\tilde{\eta}_{it} = u_{it} < c_{it} - g y_{i,t-1} - x'_{it} b$ for some random variable u_{it} that can depend on y_{i0}, x_{it}, c_{it}, and η_{it}.

With this choice, we have $\tilde{y}_{it} = y_{it}$, $\tilde{c}_{it} = c_{it}$ and $\tilde{d}_{it} = 1\{\tilde{y}^*_{it} \geq \tilde{c}_{it}\} = d_{it}$. As before (see Theorem 2.1), we want $P\{\tilde{\eta}_{i1} \leq \tau | y_{i0}, \mathbf{x}_i\} = P\{\tilde{\eta}_{i2} \leq \tau | y_{i0}, \mathbf{x}_i\}$ (i.e., we want Assumption 6 to be satisfied by $\tilde{\eta}_{i1} = \tilde{\epsilon}_{i1} + \tilde{\alpha}_i$ and $\tilde{\eta}_{i2} = \tilde{\epsilon}_{i2} + \tilde{\alpha}_i$. The proof then follows the proof of Theorem 2.1. □

A.3.7 Proof of Theorem 4.2

The proof is similar to Theorem 4.1 and therefore is omitted.

A.3.8 Proof of Theorem 4.3

For any $(g, b) \in \Theta_{I,4}$, let $\tilde{y}^*_{it} = g \tilde{y}^*_{i,t-1} + x'_{it} b + \tilde{\alpha}_i + \tilde{\epsilon}_{it}$, $\tilde{\epsilon}_{i1} = 0$, $\tilde{c}_{it} = c_{it}$, and $\tilde{y}^*_{i0} = y^*_{i0}$. Let's define $\tilde{\alpha}_i$ and $\tilde{\epsilon}_{i2}$ in a following way:

- If $y_{i1}^* \geq c_{i1}$, $y_{i2}^* \geq c_{i2}$: $\tilde{\alpha}_i = \alpha_i + \epsilon_{i1} + (\gamma - g)y_{i0}^* + x_{i1}'(\beta - b)$; and $\tilde{\epsilon}_{i2} = \Delta x_i'(\beta - b) + \epsilon_{i2} - \epsilon_{i1} + (\gamma - g)(\alpha_i + \gamma y_{i0}^* + x_{i1}'\beta + \epsilon_{i1} - y_{i0}^*)$.

- If $y_{i1}^* < c_{i1}$, $y_{i2}^* \geq c_{i2}$: $\tilde{\alpha}_i = u_i$, where $u_i < c_{i1} - g y_{i0}^* - x_{i1}'b$; and $\tilde{\epsilon}_{i2} = \alpha_i + \epsilon_{i2} - \tilde{\alpha}_i + (\gamma y_{i1}^* - g\tilde{y}_{i1}^*) + x_{i2}'(\beta - b)$.

- If $y_{i1}^* \geq c_{i1}$, $y_{i2}^* < c_{i2}$: $\tilde{\alpha}_i = \alpha_i + (\gamma - g)y_{i0}^* + x_{i1}'(\beta - b) + \epsilon_{i1}$; and $\tilde{\epsilon}_{i2} = w_i$, where $w_i < c_{i2} - \tilde{\alpha}_i - g y_{i1}^* - x_{2i}'b$.

- If $y_{i1}^* < c_{i1}$, $y_{i2}^* < c_{i2}$: $\tilde{\alpha}_i = u_i$, where $u_i < c_{i1} - g y_{i0}^* - x_{i1}'b$; and $\tilde{\epsilon}_{i2} = w_i$, where $w_i < c_{i2} - \tilde{\alpha}_i - g\tilde{y}_{i1}^* - x_{2i}'b$.

Then the lower and upper bounds on the conditional distribution of $\tilde{\epsilon}_{i2}|\mathbf{x}_i, \tilde{y}_{i0}^*$, obtained by varying distributions of u_i and w_i within their bounds, are $P\{y_{i2}^U - (1+g)y_{i1}^L + g y_{i0}^* - \Delta x_i'b \leq \tau | \mathbf{x}_i, y_{i0}^*\}$ and $P\{y_{i2}^L - (1+g)y_{i1}^U + g y_{i0}^* - \Delta x_i'b \leq \tau | \mathbf{x}_i, y_{i0}^*\}$, correspondingly. Intersecting these bounds over the support of $\mathbf{x}_i$ allows us to find the distribution of $\tilde{\epsilon}_{i2}$ that is independent of $\mathbf{x}_i$ and $\tilde{y}_{i0}^*$, so that Assumption 9 is satisfied. □

A.3.9 Proof of Theorem 4.4

The proof closely follows the proof of Theorem 3.1 with the following choices: $\tilde{c}_{it} = c_{it}$, $\tilde{y}_{it}^* = g_t\tilde{\alpha}_i + \tilde{\epsilon}_{it} + x_{it}'b$, where $g_1 = 1$, $g_2 = g$, $\tilde{\epsilon}_{i1} = 0$ and

- If $y_{i1}^* \geq c_{i1}$, $y_{i2}^* \geq c_{i2}$: $\tilde{\alpha}_i = \alpha_i + x_{i1}'(\beta - b) + \epsilon_{i1}$; and $\tilde{\epsilon}_{i2} = (\gamma - g)\alpha_i + (x_{i2} - g x_{i1})'(\beta - b) + \epsilon_{i2} - g\epsilon_{i1}$.

- If $y_{i1}^* < c_{i1}$, $y_{i2}^* \geq c_{i2}$: $\tilde{\alpha}_i = u_i$, where $u_i < c_{i1} - x_{i1}'b$; and $\tilde{\epsilon}_{i2} = \gamma\alpha_i - g\tilde{\alpha}_i + x_{i2}'(\beta - b) + \epsilon_{i2}$.

- If $y_{i1}^* \geq c_{i1}$, $y_{i2}^* < c_{i2}$: $\tilde{\alpha}_i = \alpha_i + x_{i1}'(\beta - b) + \epsilon_{i1}$; and $\tilde{\epsilon}_{i2} = w_i$, where $w_i < c_{i2} - g\tilde{\alpha}_i - x_{2i}'b$.

- If $y_{i1}^* < c_{i1}$, $y_{i2}^* < c_{i2}$: $\tilde{\alpha}_i = u_i$, where $u_i < c_{i1} - x_{i1}'b$; and $\tilde{\epsilon}_{i2} = w_i$, where $w_i < c_{i2} - g\tilde{\alpha}_i - x_{2i}'b$.

Then the upper and lower bounds on the conditional distribution of $\tilde{\epsilon}_{i2}/g|\mathbf{x}_i$,obtained by varying distributions of u_i and w_i within their bounds, are $P\{y_{i2}^L/g - y_{i1}^U - (x_{i2}'/g -$

$x_{i1})'b \leq \tau | \mathbf{x}_i\}$ and $P\{y_{i2}^U/g - y_{i1}^L - (x_{i2}'/g - x_{i1})'b \leq \tau | \mathbf{x}_i\}$, correspondingly. Intersecting these bounds over the support of $\mathbf{x}_i$ allows us to find the distribution of $\tilde{\epsilon}_{i2}$ that is independent of $\mathbf{x}_i$. □

A.3.10 Proof of Theorem 5.1

Assumption 10 together with Assumption 11 and the weak law of large numbers implies that $\hat{P}\{\mathbf{x}_i = \mathbf{x}\}$ converges in probability to $P\{\mathbf{x}_i = \mathbf{x}\}$ uniformly over $\mathbf{x} \in \mathcal{X}$. Next, consider the following empirical processes indexed by $(\tau, \mathbf{x})$:

$$\nu_{1,n}(\tau, \mathbf{x}; \beta) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(1\{y_{i1}^U - x_{i1}'\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i2}^L - x_{i2}'\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\}\right) - D_1(\tau, \mathbf{x}; \beta)\right]$$

$$\nu_{2,n}(\tau, \mathbf{x}; \beta) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(1\{y_{i2}^U - x_{i2}'\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\} - 1\{y_{i1}^L - x_{i1}'\beta \leq \tau, \mathbf{x}_i = \mathbf{x}\}\right) - D_2(\tau, \mathbf{x}; \beta)\right]$$

The class of functions $1\{z_i \leq \tau, \mathbf{x}_i = \mathbf{x}\}$ is a VC-class, and therefore is P-Donsker. That is,

$$(\nu_{1,n}(\cdot, \cdot; \beta), \nu_{2,n}(\cdot, \cdot; \beta)) \Rightarrow (G_1(\cdot, \cdot; \beta), G_2(\cdot, \cdot; \beta))$$

Under the null hypothesis, $D_1(\cdot, \cdot; \beta), D_2(\cdot, \cdot; \beta) \leq 0$. Together with the weak convergence result above and the continuous mapping theorem for empirical processes, this implies the asymptotic first-order dominance result for the distribution of $T_n^S(\beta)$. □

A.3.11 Proof of Theorem 5.2

The proof of this result closely follows the proof of Theorem 5.1 and therefore is omitted here. □

A.3.12 Proof of Theorem A.1

We construct the following random variables: $d_{i2}^L = I\{y_{i2}^L \leq \tau_{i2}\}$ and $d_{i1}^U = I\{y_{i1}^U \leq \tau_{i1}\}$, where τ_{i1} and τ_{i2} satisfy conditions of Theorem A.1. As before, let $\eta_{it} = \alpha_i + \epsilon_{it}$.

Then $E[d_{i2}^L|\mathbf{x}_i,\tau_{i2}] = P\{y_{i2}^L \leq \tau_{i2}|\mathbf{x}_i,\tau_{i2}\} = 1 - P\{y_{i2}^L > \tau_{i2}|\mathbf{x}_i,\tau_{i2}\} = 1 - P\{x_{i2}'\beta + \eta_{i2} > c_{i2}, x_{i2}'\beta + \eta_{i2} > \tau_{i2}|\mathbf{x}_i,\tau_{i2}\} = 1 - P\{x_{i2}'\beta + \eta_{i2} > \max\{c_{i2},\tau_{i2}\}|x_{i2},\tau_{i2}\} = P\{\eta_{i2} < \tau_{i2} - x_{i2}'\beta|\mathbf{x}_i,\tau_{i2}\}$ Here the last equality follows from Assumption 13.

Similarly, $E[d_{i1}^U|\mathbf{x}_i,\tau_{i1}] = P\{\max\{x_{i1}'\beta+\eta_{i1}, c_{i1}\} \leq \tau_{i1}|\mathbf{x}_i,\tau_{i1}\} = P\{x_{i1}'\beta+\eta_{i1} \leq \tau_{i1}, c_{i1} \leq \tau_{i1}|\mathbf{x}_i,\tau_{i1}\} = P\{\eta_{i1} \leq \tau_{i1} - x_{i1}'\beta|\mathbf{x}_i,\tau_{i1}\}$.

Finally, taking into account that $\eta_{i1} = \epsilon_{i1} + \alpha_i$ and $\eta_{i2} = \epsilon_{i2} + \alpha_i$ are identically distributed conditional on $\mathbf{x}_i$, we have: $E[d_{i2}^L|\mathbf{x}_i,\tau_{i2}] = F_\eta(\tau_{i2} - x_{i2}'\beta|\mathbf{x}_i)$ and $E[d_{i1}^U|\mathbf{x}_i,\tau_{i1}] = F_\eta(\tau_{i1} - x_{i1}'\beta|\mathbf{x}_i)$, where $F_\eta(\cdot|\mathbf{x}_i)$ is a c.d.f. of η_{it} conditional on $\mathbf{x}_i$. Now, taking into account that F_η is a strictly monotone function for any value of $\mathbf{x}_i$, we have

$$E[d_{i1}^U - d_{i2}^L|\mathbf{x}_i,\tau_{i1},\tau_{i2}] > 0 \text{ if and only if } \Delta\tau_i < \Delta x_i'\beta, \tag{A.7}$$

where $\Delta\tau_i = \tau_{i2} - \tau_{i1}$ and $\Delta x_i = x_{i2} - x_{i1}$. Consequently, point identification follows from identical arguments used in Khan and Tamer (2007). □

A.3.13 Proof of Theorem A.2

Part 1. Suppose that $b \in B$ is such that $b_k \neq \beta_k$. Then assumption Assumption 12 implies that $\Delta x_i'(b-\beta)$ and $\Delta x_j'(b-\beta)$ are unbounded on the support of $\mathbf{x}_i$. Therefore, for any $0 < \delta < 1$ and any τ we can find such values of $\mathbf{x}_i$ and $\mathbf{x}_j$ that $F_{\Delta\epsilon}(\tau + \Delta x_i'(b-\beta)|\mathbf{x}_i) - F_{\Delta\epsilon}(\tau + \Delta x_j'(b-\beta)|\mathbf{x}_j) > \delta$. Let $q < \delta < 1$. Then we have $F_{\Delta\epsilon}(\tau + \Delta x_i'(b-\beta)|\mathbf{x}_i) - F_{\Delta\epsilon}(\tau + \Delta x_j'(b-\beta)|\mathbf{x}_j) > q$ for some $\mathbf{x}_i$ and $\mathbf{x}_j$, which is a contradiction to Assumption 14(i). Therefore, β is identified relative to b.

Part 2. Suppose now that $b \in B$ is such that $b_k = \beta_k$ but $b \neq \beta$. Assumption 12 ensures that there exist some $\gamma_2 < \gamma_1$ such that the sets $\overline{\mathcal{X}}_{\gamma_1} = \{\mathbf{x}_{i,-k}:$ such that $\Delta x_i'(b-\beta) = \Delta x_{i,-k}'(b_{-k} - \beta_{-k}) > \gamma_1\}$ and $\underline{\mathcal{X}}_{\gamma_2} = \{\mathbf{x}_{j,-k}:$ such that $\Delta x_j'(b-\beta) = \Delta x_{j,-k}'(b_{-k}-\beta_{-k}) < \gamma_2\}$ are nonempty. Then there exist $\rho > 0$ and $\tilde{\tau}$ such that $H(x_{i,-k}, x_{j,-k}) \equiv F_{\Delta\epsilon}(\tilde{\tau} + \Delta x_i'(b-\beta)|\mathbf{x}_i) - F_{\Delta\epsilon}(\tilde{\tau} + \Delta x_j'(b-\beta)|\mathbf{x}_j) > \rho$ on $\mathcal{X}_{\gamma_1,\gamma_2} = \overline{\mathcal{X}}_{\gamma_1} \times \underline{\mathcal{X}}_{\gamma_2}$. Hence, the left-hand side of (A.1) is bounded away from zero for $\tau = \tilde{\tau}$ on $\mathcal{X}_{\gamma_1,\gamma_2}$ for all values of $\mathbf{x}_{i,k}$ and $\mathbf{x}_{j,k}$ in the support. On the other hand, Assumption 14(ii) implies that the right-hand side of (A.1) can be made less than any $\rho > 0$ with a proper choice of $\mathbf{x}_{i,k}$ and $\mathbf{x}_{j,k}$. Therefore, β is identified relative to any $b \neq \beta$, so that $B_I = \{\beta\}$. □

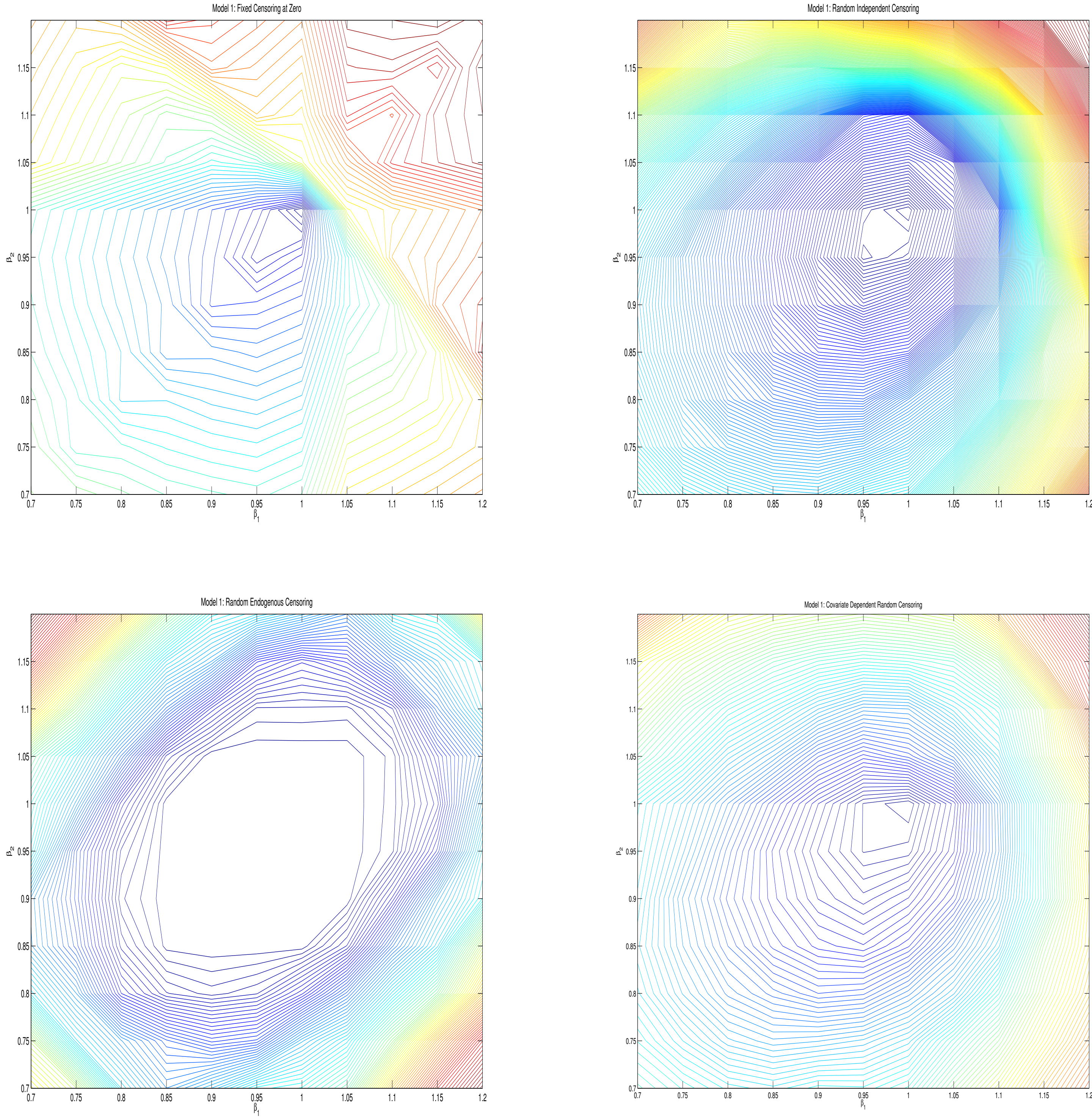

Figure 1: Fixed and Random Independent Censoring (top) Endogenous censoring and covariate dependent censoring (bottom)

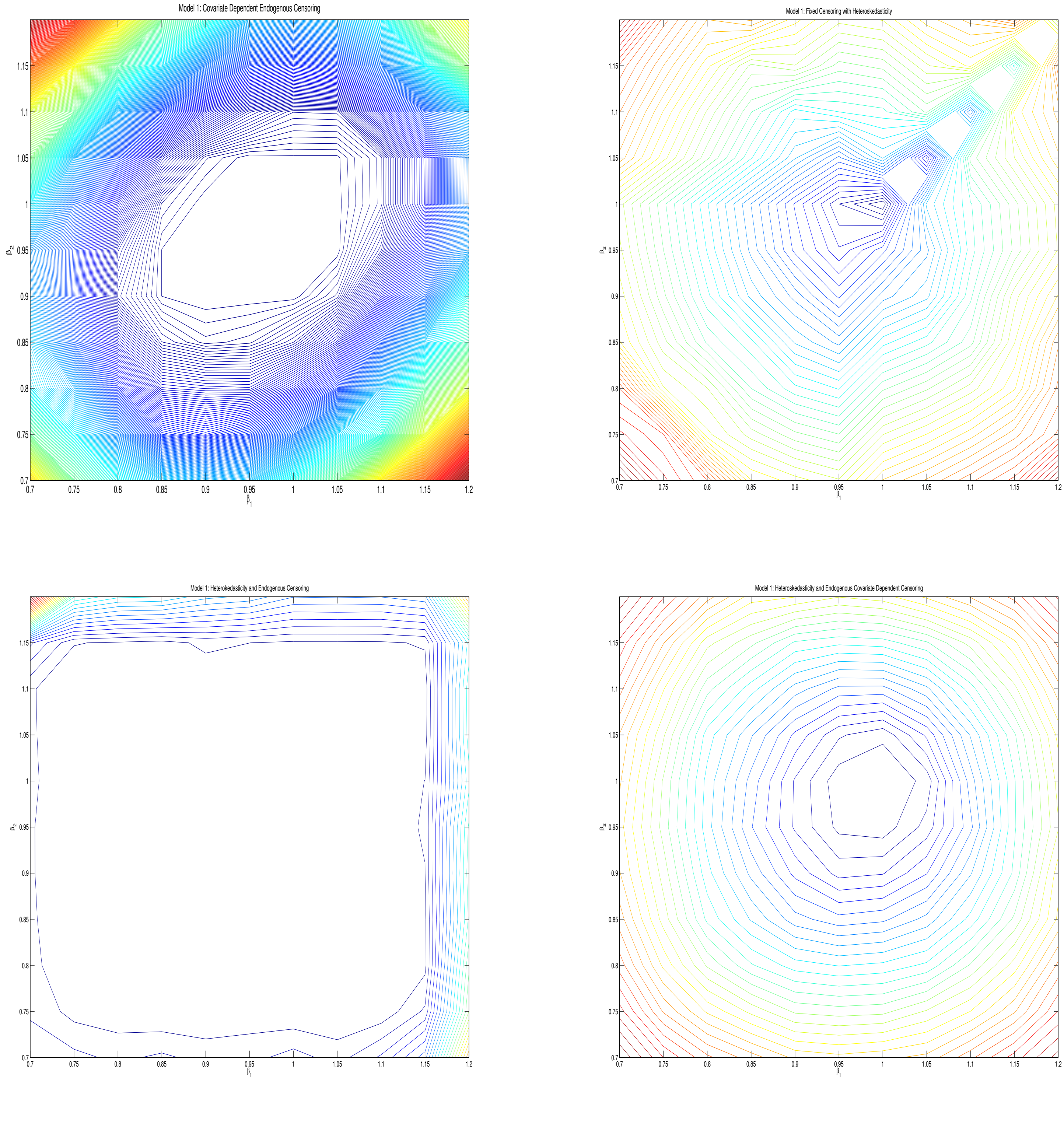

Figure 2: Model 1: Covariate dependent endogenous censoring, fixed censoring with heteroskedasticity (top) Heteroskedastic endogenous censoring and heteroskedastic covariate dependent censoring (bottom)

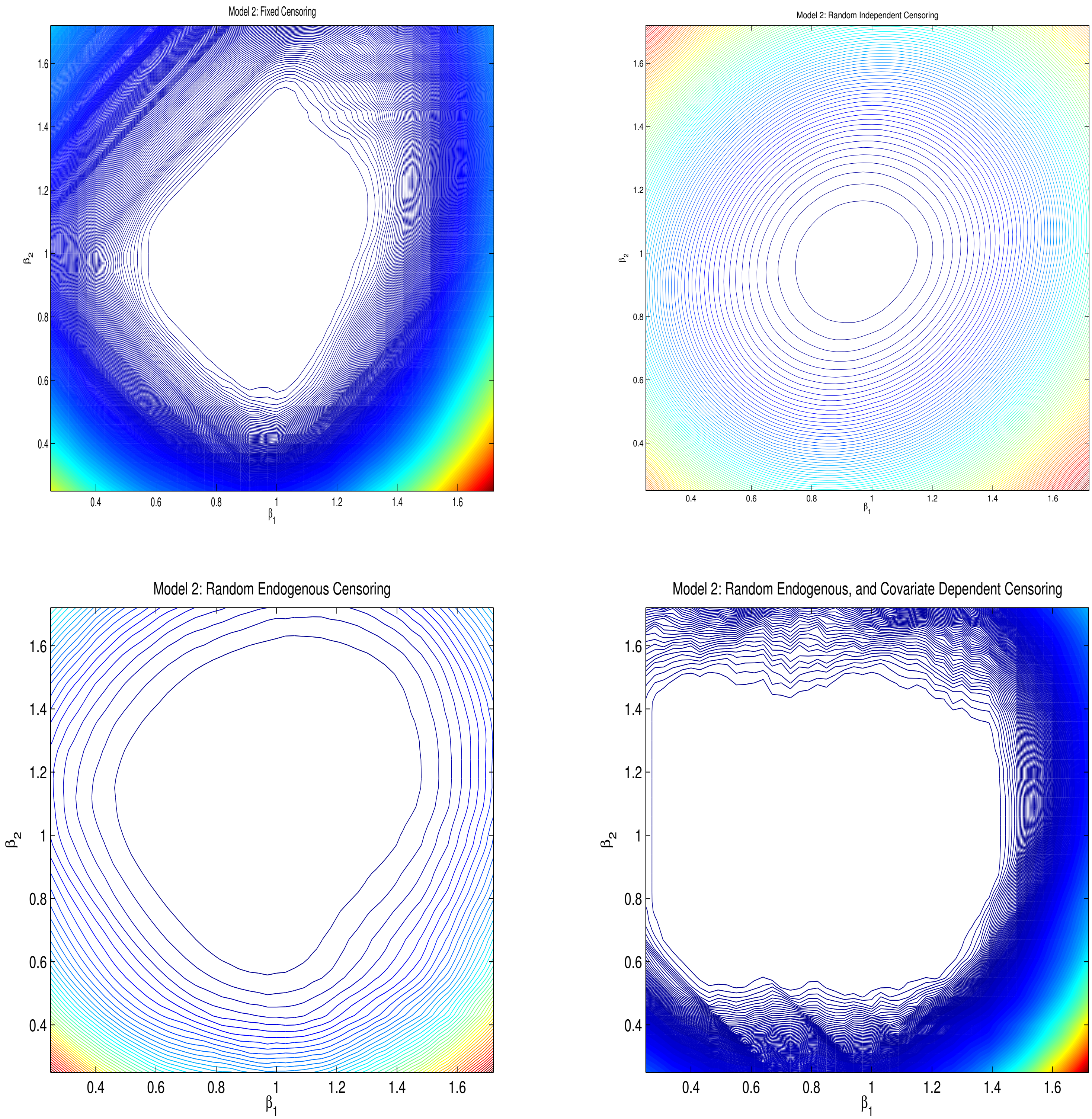

Figure 3: Model 2: Fixed censoring and random independent censoring (top) endogenous censoring and endogenous covariate dependent censoring (bottom)

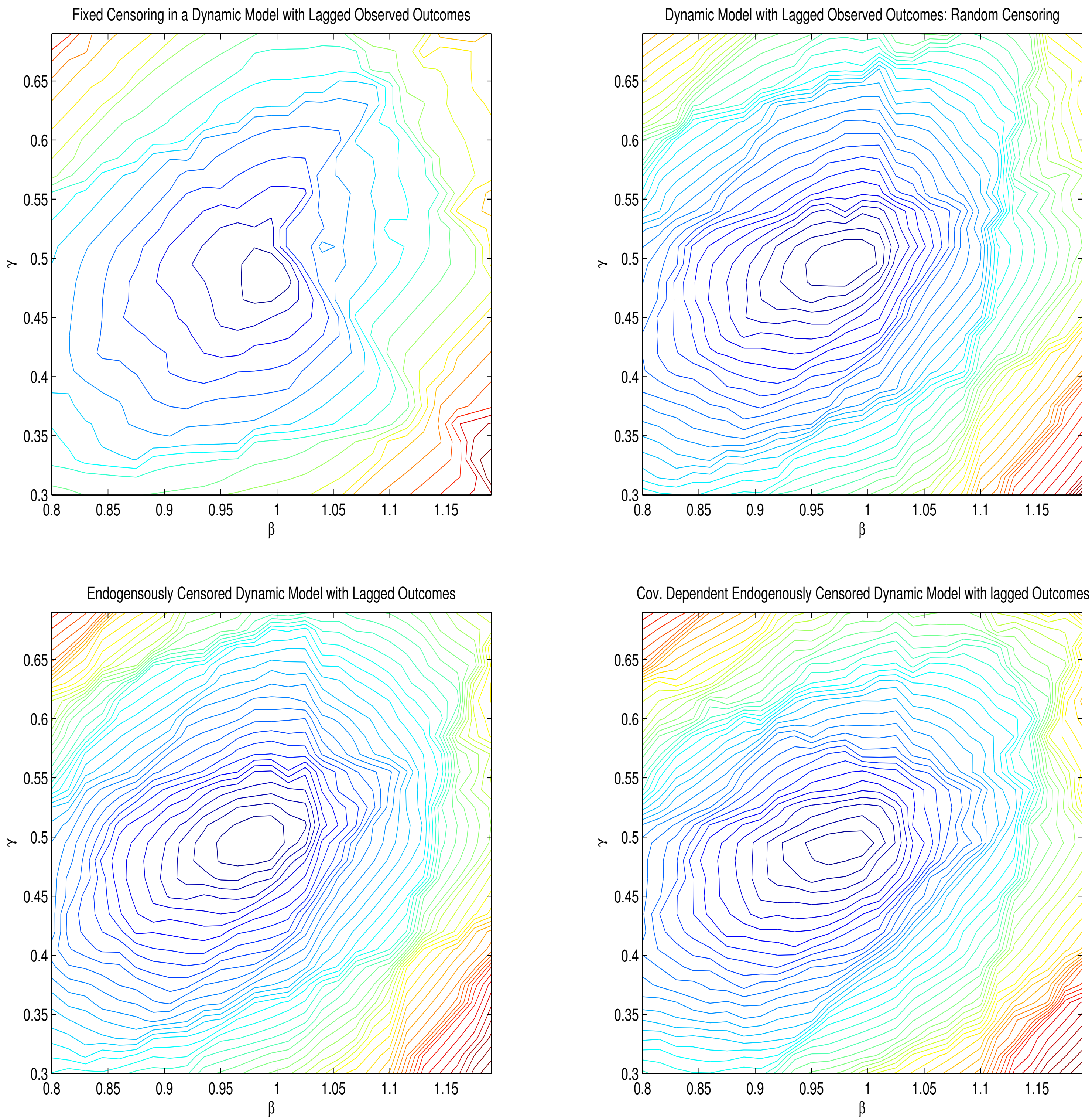

Figure 4: Dynamic Model with Lagged Outcomes: Fixed censoring and random independent censoring (Top) Endogenous and covariate dependent censoring (Bottom)

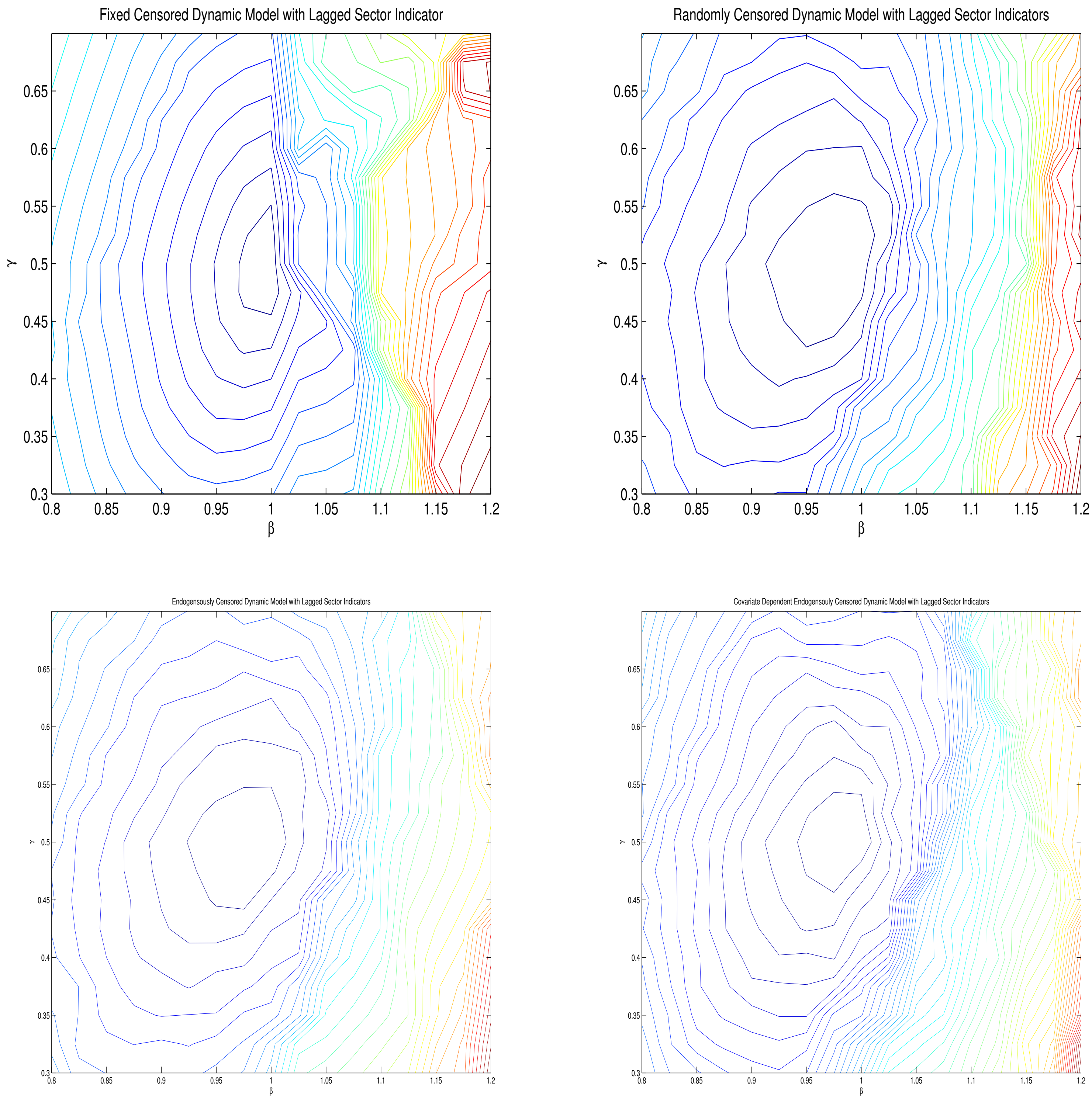

Figure 5: Dynamic Model with Lagged Sector Indicators: Fixed Censoring and Random Independent Censoring (Top) Endogenous Censoring (Bottom)