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
University of Western Ontario
London, ON - Canada
mponomar@uwo.ca

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

April 18, 2011

Abstract

This paper analyzes the identification question in censored panel data models, where the censoring can depend on both observable and unobservable variables in arbitrary ways. Under some general conditions, we derive the tightest sets on the parameter of interest. These sets (which can be singletons) represent the limit of what one can learn about the parameter of interest given the model and the data in that every parameter that belongs to these sets is *observationally equivalent* to the true parameter. We consider two separate sets of assumptions, motivated by the previous literature, each controlling for unobserved heterogeneity with an individual specific (fixed) effect. The first imposes a stationarity assumption on the unobserved disturbance terms, along the lines of Manski (1987), and Honoré (1993). The second is a *nonstationary* model that imposes a conditional independence assumption. For both models, we provide sufficient conditions for these models to point identify the parameters. Since our identified sets are defined through parameters that obey first order dominance, we outline easily implementable approaches to build confidence regions based on recent advances in Linton et.al.(2010) on bootstrapping tests of stochastic dominance. We also extend our results to dynamic versions of the censored panel models in which we consider lagged observed, latent dependent variables and lagged censoring indicator variables as regressors.

*We thank Marcelo Ochoa for research assistance on this project, and for the National Science Foundation for research support.

†Corresponding author.

1 Introduction

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

$$y_{it}^* = \alpha_i + x_{it}'\beta + \epsilon_{it}, \quad t = 1, \ldots T$$

where α_i is an individual specific and time-independent random variable -or fixed effect- that is allowed to be correlated with both $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{iT})$ and $\epsilon_i^T = (\epsilon_{i1}, \ldots, \epsilon_{iT})$. Complications arise because, first, the outcome variable, y_{it}^*, is only observed when it is larger than a random variable c_{it}[1]; and second, $\mathbf{c}_i^T = (c_{i1}, \ldots, c_{iT})$ is allowed to depend on ϵ_i^T conditional on $\mathbf{x}_i^T$, i.e., we

$$\begin{array}{ll} \text{observe for } i: & (\max(y_{it}^*, c_{it}),\ 1[y_{it} \geq c_{it}],\ x_{it}) \quad t = 1, \ldots, T \\ \text{with} & \epsilon_i^T \not\perp \mathbf{c}_i^T | \mathbf{x}_i^T \end{array}$$

The presence of this *endogenous* censoring represents a challenge for existing methods[2] that are used for correcting for censoring since these methods usually assume that c 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[3]). This enlarges the set of models that are covered to include competing risks and switching regression like models that are important in applied economics.

Generally, point identification conditions in nonlinear panel data models 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

[1] c_{it} itself is only observed when it exceeds y_{it}^*.

[2] This is especially the case when T is finite, as we assume throughout this paper.

[3] 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.

the fact that though point identification might not be possible under weaker assumptions, these models do contain information about the parameter of interest. 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 the parameter of interest given the observed data.

The challenge in this approach to identification analysis is to exhaust all the information in the data and the model: that is, find the *tightest* set. 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 identified sets* for β: there is no more information that the data contain about β 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 as opposed to the complementary approach based on point identification in which one looks for a model (a set of assumptions) that guarantee point identification.

There are a set of recent papers that deal with various nonlinearities in models with (short T) panels. See for example the work of Bester and Hansen (2009), Bonhomme (2010), Chernozhukov, Fernandez-Val, Hahn, and Newey (2010), Evdokimov (2010), Graham and Powell (2009) and Hoderlein and White (2009). See also the survey in Arellano and Honoré (2001). We study the linear model above under censoring where the outcomes (y_{it}^*, c_{it}) are partially observed. 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. Censoring can also be a result of mechanical considerations such as top-coding, and there, typically, the censoring is fixed (and hence exogenous). Our approach to inference will cover both cases, endogenous and exogenous, naturally. 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. Both these models are useful in empirical settings.

Generally, missing or interval outcome models were considered in a nonparametric setup in the partial identification literature. Manski and Tamer (2002) considered inference on the slope vector in a linear model with interval outcomes using a partial identification approach. With panel data, Honoré and Tamer (2006) and Chernozhukov, Fernandez-Val, Hahn, and Newey (2010) have considered bounds on parameters of interest in some interesting nonlinear 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), and weaken them to allow for arbitrary censoring that can be correlated with both the outcomes and the covariates, while allowing for 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. Similar trade-off is shown, for example, by Rosen (2009) for quantile panel data models with fixed effects and small T. In particular, under a conditional median independence assumption on ϵ_{it}, Rosen (2009) 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.

We employ two sets of (relatively weak) assumptions that allow us to non-trivially identify the parameters in the censored linear regression model. The first set of assumptions (**Model 1**) uses *stationarity* on the distribution of ϵ^T, 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 assumptions (and support conditions). See also Honoré (1993) and Chernozhukov, Fernandez-Val, Hahn, and Newey (2010). We show that a particularly constructed sets of *moment inequalities* characterize the sharp set, B_I on β. The proof basically shows that *any* parameter $b \in B_I$ is observationally equivalent to β given the maintained assumptions and the data: given $b \in B_I$ one can construct an error distribution that obeys stationarity, and generates the observed data.

The second set of assumptions (**Model 2**) relaxes stationarity, but maintains independence between ϵ^T and $\mathbf{x}^T$. This non-stationary model allows for *arbitrary* correlation in the errors across time periods. Here also, another set of conditional moment inequalities is shown to sharply characterize the identified set. Using the structure of those inequalities, one can obtain conditions under which the model contains no information.

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.

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 CDFs, and hence conducting inference is similar to testing whether one CDF stochastically dominates another; and so, one way to do inference is to adapt some recently developed methods from the stochastic dominance literature to our setup.

The next section defines the model above under stationarity, and Section 3 gives sufficient conditions for sharp identification and provide a consistent estimator under these conditions. In this section we also provide conditions under which the vector of parameter in the stationary model is point identified. In Section 4, we replace stationarity with an independence (but not necessary stationary) assumption and derive the sharp set under these conditions. Section 5 proposes an inference procedure for the parameters of interest that is based on bootstrapping a particular stochastic dominance test statistic (Linton, Song, and Whang (2010)). Also, for the point-identified stationary case we propose a simple rank-based estimator of β. Section 6 modifies our approach to identify parameters in dynamic models, analogous to those considered in (Hu (2002)). Section 7 provides numerical evidence on the size of the identified set in some examples and section 8 concludes.

2 Censored Panel Data Model

To illustrate the identification approach taken in this paper we first introduce a censored panel data model. We will characterize the model within the linear latent dependent framework. Here the latent dependent variable associated with the sector whose parameters we wish to identify is denoted by:

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

where $i = 1, 2, ...N,\ t = 1, 2$. α_i is an unobserved individual specific "fixed" effect, and we assume the unobserved disturbance terms $\epsilon_{i1}, \epsilon_{i2}$ are strictly stationary given $\mathbf{x}_i \equiv (x_{i1}, x_{i2})$ and α_i in Model 1, and arbitrary correlated but independent of $\mathbf{x}_i$ in Model 2. 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. The number of time periods T is set to 2 w.l.o.g. We are only emphasizing that the number of time periods T is small relative to the number of cross-sectional units N which is assumed to be large and going to $+\infty$.

The econometrician observes $v_{it} \equiv \max(y_{it}^*, c_{it})$, where, for example, c_{it} denotes the wage offered in a different sector, and the indicator d_{it}, which denotes which sector the wage is drawn from. Note we impose no structure on c_{it} here, regarding features of its distribution as nuisance parameters for now. This will enable the framework discussed below to also estimate the statistical analog of the Roy model- the *competing risks model*, which further nests randomly censored panel data models. Hence the models studied here generalize existing work on censored panel data models e.g. Honoré (1992), Honoré, Khan, and Powell (2002)- in the sense that the censoring variable can be random and more importantly can be correlated with $\mathbf{x}_i, \alpha_i, \epsilon_i$.

The question that is at hand is: how do we map assumptions made on the joint distribution of $\epsilon_{i1}, \epsilon_{i2}|\mathbf{x}_i, \alpha_i$ 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 $\epsilon|x$ 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 finite bounds on β even in the absence of censoring. This is so because we do not place any restrictions on the correlation structure of the vector $(\epsilon_{i1}, \epsilon_{i2})$. See the recent work in Rosen (2009) where this point was made for panel models with no censoring. So, then, we know that with censoring, stronger assumptions are needed to obtain non-trivial bounds on β. Under the stationarity assumptions of Model 1 and the independence assumptions of Model 2, we will show below that a set of bounds on particularly defined and observed conditional distribution functions characterize the identified set. One of the main contributions of this paper is to show that the bounds we derive 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 (2008).

So, under the censoring mechanism that we consider with panel data and endogenous censoring, we formally show that our bounds exhaust all the information in the model. we will start with stationarity.

3 Identification with Stationarity

In this section we propose an inference procedure under the assumption of conditional stationarity on the disturbance terms.

Model 1: $\epsilon_{i1} + \alpha_i$ has the same distribution as $\epsilon_{i2} + \alpha_i$ 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. Obviously, the censoring complicates the problem and so below, we provide the information that the *observed data* contains about β *under Model 1.*

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 that ϵ_{i1} and ϵ_{i2} have the same distribution conditional on the vector of covariates $x_i = (x_{i1}, x_{i2})$ and the fixed effect. In each period, a researcher observes only (v_{it}, d_{it}, x_{it}), where $v_{it} = \max\{y_{i1}^*, c_{it}\}$ and $d_{it} = 1\{y_{i1}^* > c_{it}\}$. We start with constructing a sharp identified set for β without placing any restrictions on censoring variables c_{it}. We define the following variables:

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

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

$$y_{it}^L \le y_{it}^* = x_{it}'\beta + \alpha_i + \epsilon_{it} \le y_{it}^U \tag{3.1}$$

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 (3.1) naturally imply that the parameter β satisfies the following set of *conditional moment inequalities* for any τ and any x_i:

$$\begin{aligned} P\{y_{i1}^U - x_{i1}'\beta \le \tau | \mathbf{x}_i\} \le P\{y_{i2}^L - x_{i2}'\beta \le \tau | \mathbf{x}_i\} \\ P\{y_{i2}^U - x_{i2}'\beta \le \tau | \mathbf{x}_i\} \le P\{y_{i1}^L - x_{i1}'\beta \le \tau | \mathbf{x}_i\} \end{aligned} \tag{3.2}$$

We define the *identified set* B_I as

$$B_I = \{b \in B: \text{ for any } \tau \in R \text{ and } \mathbf{x}_i, \text{(3.2) holds with } \beta = b\} \tag{3.3}$$

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.* This will be shown in the next Theorem which is the main result in this section.

Theorem 3.1 *Any $b \in B_I$ is observationally equivalent to β and so B_I is the sharp set.*

Proof: See Appendix.

Remark 3.1 *The set B_I above is non empty since under well specification, the true parameter β belongs to the set. It is easy to see that the set B_I is convex. Also, the stationarity assumptions although is restrictive it does allow for correlation between ϵ_1 and ϵ_2, and more importantly also allows for cross sectional heteroskedasticity.*

Remark 3.2 *We 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 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.*

An immediate Corollary to the above Theorem follows.

Corollary 3.1 *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 any* 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). Note also that the inferential strategy above is based directly on the stationarity assumption and it does not require any explicit differencing to get rid of the fixed effects.
Next, we analyze the above inequalities and provide a sufficient condition under which the set B_I shrinks to a point, i.e., we achieve point identification.

3.1 Attaining Point Identification Under Stationarity

Here we establish a sufficient condition for point identification. Our sufficient condition is that for some $\mathbf{x}_i$,α_i where $\Delta\mathbf{x}_i = x_{i1} - x_{i2}$ satisfies the usual full rank condition, we have that $c_{it} \leq \tau_{it}$ with probability 1 for some known random variable τ_{it} that does not depend on x_i. This sufficient condition assumes basically that we have fixed or bounded support censoring[4].

Theorem 3.2 *Let $\tau_i = (\tau_{i1}, \tau_{i2})$ be independent of $\mathbf{x}_i = (x_{i1}, x_{i2})$ and, α_i. Assume that the random variable $\Delta\mathbf{x}_i$ has full rank on Ξ where*

$$\Xi = \{\mathbf{x}_i : \quad P(c_{it} \leq \tau_{it} \,|\mathbf{x}_i, \alpha_i) = 1\} \tag{3.4}$$

and $\Delta\mathbf{x}_i = x_{i1} - x_{i2}$. Then, $B_I = \{\beta\}$ and so β is point identified.

[4]In certain settings such as independent or conditional independent censoring, this condition is not necessary.

Proof: See Appendix.

Under the support condition above, one can consistently estimate β using a rank-based estimator. This estimation approach is outlined in Section 5 below.

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 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 ϵ_1 and ϵ_2 to be arbitrarily correlated.

4 Non- Stationary Model

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 marginals of ϵ's do not allow us to provide *any* restrictions on β, i.e., the *sharp* set is the *trivial* set- i.e. the original parameter space. 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:

Model 2: The vector $(\epsilon_{i1}, \epsilon_{i2})$ is independent of $\mathbf{x}_i = (x_{i1}, x_{i2})$.

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, for the analysis in this section to go through, all we need is for the difference $\Delta\epsilon = \epsilon_1 - \epsilon_2$ to be independent of the vector $\mathbf{x}$.

As before, we consider the same two-period panel data model:

$$y^*_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it}, \text{ where } t = 1, 2.$$

Both y^*_{i1} and y^*_{i2} are only partially observed, and both ε_{i1} and ε_{i2} are unobserved. Here we assume that ε_{i1} and ε_{i2} are distributed independently of the observed vector $x_i = (x_{i1}, x_{i2})$, but we do not require the errors to be distributed independently of fixed effects α_i's. To fit into our framework, one observes $v_{it} = \max\{y_{i1}^*, c_{it}\}$ and $d_{it} = 1\{y_{i1}^* > c_{it}\}$. 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. As in the previous section, we define the following variables:

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

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

$$
y_{it}^L \leq y_{it}^* \leq y_{it}^U
$$

Note since this holds for the pair of observations $t = 1, 2$ and thus will imply the following inequalities that do not contain α_i:

$$
y_{i2}^L - y_{i1}^U \leq \triangle\mathbf{x}_i'\beta + \triangle\varepsilon_i \leq y_{i2}^U - y_{i1}^L
$$

where $\triangle\mathbf{x}_i = x_{i2} - x_{i1}$ and $\triangle\varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1}$. Since we assume that ε is independent of $\mathbf{x}_i$ this means that $\Delta\varepsilon$ 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 4.1 *For any b in the parameter set B, define*

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

and

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

Then the set

$$
B_I = \{b \in B : \textit{ for all } \mathbf{x}_i, \mathbf{x}_j \textit{ and } \tau \ LB(\tau, \mathbf{x}_i, b) \leq UB(\tau, \mathbf{x}_j, b)\} \tag{4.1}
$$

is the sharp identified set for β.

Proof: See Appendix.

Again, here, parameters of regressors that do not change through time cannot be identified from the fixed effect. The size of the identified set B_I 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 4.2 *For $t = 1, 2$ define $p_t(\mathbf{x}_i) = 1 - P(d_{it} = 1|\mathbf{x}_i) = P\{y_{it} < c_{it}|\mathbf{x}_i\}$. If for all $\mathbf{x}_i$ and $\mathbf{x}_j$ we have $p_1(\mathbf{x}_i) + p_2(\mathbf{x}_j) \geq 1$, then any $b \in B$ is observationally equivalent to β, so that $B_I = B$.*

The above is an interesting result that basically says that even under the independence assumption, Model 2 contains no restrictions if there is a lot of censoring. Basically, the result requires that censoring be higher than 50%.

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

4.1 Sufficient Conditions for Point Identification

It is interesting to see under what conditions β is point identified. As we already noted above, if no censoring occurs for a subset of the support of $\mathbf{x}_i$ such that the corresponding subset of the support of $\triangle\mathbf{x}_i$ 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 x_i. We start by defining $p(\mathbf{x}_i) = P\{y_{i1}^* > c_{i1}, y_{i2}^* > c_{i2}|\mathbf{x}_i\}$. Then, given UB in (A.3), we have

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

Similarly, given (A.2) above,

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

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

$$F_{\triangle\varepsilon}(\tau + \triangle\mathbf{x}_i'(b-\beta)) - F_{\triangle\varepsilon}(\tau + \triangle\mathbf{x}_j'(b-\beta)) \leq 2 - p(\mathbf{x}_i) - p(\mathbf{x}_j) \quad (4.2)$$

for any τ, $\mathbf{x}_i$ and $\mathbf{x}_j$, where $F_{\triangle\varepsilon}(\cdot)$ denotes the conditional CDF of $\triangle\varepsilon_i$. This motivates the following *sufficient conditions* for point identification of β.

A1 **Large support:** (i) Conditional on all other components (denoted by subscript $-k$), the distribution of k^{th} component of vector $\triangle\mathbf{x}_i$ is absolutely continuous on $\mathbb{R}$ with respect to Lebesgue measure, $supp(\triangle\mathbf{x}_{i,k}|\triangle\mathbf{x}_{i,-k}) = \mathbb{R}$, and $\beta_k \neq 0$. (ii) The support of $\triangle\mathbf{x}_i$ is not contained in any proper linear subspace of $\mathbb{R}^k$.

A2 **Censoring:** (i) There exists $0 < q < 1$ such that for any $\mathbf{x}_i$, $\mathbf{x}_j$ it holds that $2 - p(\mathbf{x}_i) - p(\mathbf{x}_j) < q$. (ii) For any $\mathbf{x}_{i,-k}$, $\sup_{\mathbf{x}_{i,k}\in\mathbb{R}} p(\mathbf{x}_{i,k}, \mathbf{x}_{i,-k}) = 1$.

The following theorem uses identification at infinity argument to point identify either β or its k^{th} component. Note that assumptions A1 and A2 by no means are necessary conditions for point identification.

Theorem 4.3 *Let assumptions A1 and A2(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 A2(ii) holds, then β is point identified, so that $B_I = \{\beta\}$.*

Proof: See Appendix.

The point identification result above relies on variation at infinity to shrink the set B_I to a 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 above.

To conduct statistical inference and build confidence regions here, the next section expresses the identified set as a solution to an optimization problem.

4.2 B_I as an M, U-Estimation Problem

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 an Argmin set) of a particularly defined objective function. For instance, let τ_{1i}, τ_{2i} be two iid 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 $w_i^U = y_{i2}^U - y_{i1}^L$. For any $b \in B$, define

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

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 4.4 *Assume that random variables τ_{1i} and τ_{2i} are identically continuously distributed on $(-\infty, +\infty)$ and independent of x_i and x_j. Let $B_Q = \{b : Q(b) = 0\}$. Then $B_I = B_Q = \arg\min_b Q(b)$.*

Proof: See Appendix.

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 above objective function will admit a *unique* minimum under the conditions of Theorem 4 above. 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 $Q(.)$. We do not pursue this in this paper.

4.3 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 $(\varepsilon_{i1}, \varepsilon_{i2})$ and fixed effects α_i's. Therefore, the key identifying assumption is that the vector of error terms $(\varepsilon_{i1}, \varepsilon_{i2})$ 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 $\triangle\varepsilon_i = \varepsilon_{i2} - \varepsilon_{i1}$ 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 $Med(\triangle\varepsilon_i|x_i) = 0$. In this case the identified set is also characterized by a set of conditional inequalities.

Model 3: $Med(\triangle\varepsilon_i|\mathbf{x}_i) = 0$

Theorem 4.5 *Suppose Model 3 holds. Then a sharp identified set B_I is given by $B_I = \{b \in B :$ for any $\mathbf{x}_i, \mathbf{x}_j$ $Med(y^0_{i2} - y^1_{i1}|\mathbf{x}_i) - \triangle\mathbf{x}'_i b \leq 0 \leq Med(y^1_{j2} - y^0_{j1}|\mathbf{x}_j) - \triangle\mathbf{x}'_j b\}$.*

Proof: The proof closely follows the proof of Theorem 3.1 or Theorem 4.1 and therefore is omitted.□

The Model 3 assumption is not easy to characterize in terms of restrictions on the correlation between the epsilons. On the one extreme, if ε_1 is independent and identically distributed to ε_2 (conditional on $\mathbf{x}$'s), then their difference is distributed symmetrically around 0. We conclude our discussion on nonstationarity by considering censored panel data models with time varying factor loads.

4.4 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:

$$y^*_{it} = \theta_t\alpha_i + x'_{it}\beta + \epsilon_{it} \tag{4.3}$$

where θ_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 sharp bounds on β and θ_t, assuming cross sectional homoskedasticity

We illustrate with two periods as we did before. Note here we can only identify the ratio $\theta_2/\theta_1 = \theta$, so we normalize $\theta_1 \equiv 1$. We express this as

$$y^*_{i1} = \alpha_i + x'_{i1}\beta + \epsilon_{i1} \tag{4.4}$$

$$y^*_{i2} = \theta\alpha_i + x'_{i2}\beta + \epsilon_{i2} \tag{4.5}$$

We proceed by assuming $\theta \neq 0$, and dividing both sides of the above equation by θ, yielding

$$y_{i2}^*/\theta = \alpha_i + x_{i2}'\beta/\theta + \epsilon_{i2}/\theta \tag{4.6}$$

This division immediately results in the nonstationarity of the error terms. Fortunately, the method just proposed is designed for nonstationarity. We can use the same upper and lower bounds for y_{i1}^*. For y_{i2}^*/θ we can divide the lower and upper bounds by θ if it is positive, and reverse what the lower and upper bounds are if θ is negative. As the sign of θ is unknown, this will have to be incorporated into the construction of the inequalities. So, for example we would have

$$(y_{i2}/\theta)^L = \frac{I[\theta > 0](d_{i2}v_{i2} + (1 - d_{i2})(-\infty)) + I[\theta < 0]v_{i2}}{\theta} \tag{4.7}$$

$$(y_{i2}/\theta)^U = \frac{I[\theta > 0]v_i + I[\theta < 0](d_{i2}v_{i2} + (1 - d_{i2})(-\infty))}{\theta} \tag{4.8}$$

With these bounds for the second period, we can proceed as before.

5 Inference

This section outlines approaches 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. 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 also suggest new estimators for cases where we assume that the parameters are point identified. We start with a general method based on stochastic dominance tests.

5.1 Inference Via Stochastic Dominance Test Statistic

An approach to conducting inference when the parameter is potentially partially identified is via testing whether a given value of the parameter belongs to the identified set and collecting all the parameters that cannot be rejected in a confidence like region. The structure of the identified set is one in which a conditional c.d.f. of one random variable evaluated at some parameter value is weakly smaller than the value of the conditional c.d.f. of another random variable, with both c.d.fs evaluated at the same value, τ. Since we are looking at all such values τ to determine the identified region, we are effectively determining if one random variable (conditionally) *stochastically dominates* the other. This task is somewhat similar to the approach adopted by Jun, Lee, and Shin (2011). In the setup of the distributional treatment effects, Jun, Lee, and Shin (2011) test whether it is possible to fit two distributions within bounds such that one stochastically dominates the other.

The econometrics literature has developed several tests for stochastic dominance. Fortunately, they can be adapted to conduct set inference on β for **Model 1** using the first order dominance in (3.2) which characterize the identified set of that model. Stochastic dominance tests can also be used for inference in **Model 2** using the stochastic dominance ordering in the inequalities in (4.1). We specifically employ the test in (Linton, Maasoumi, and Whang (2005)) to take into account we are interested in conditional c.d.f.s. Hence to construct a $(1-\alpha)$ confidence region for β, we will simply collect all the values of β which fail to reject the α level test in (Linton, Song, and Whang (2010)). This confidence region $\mathcal{B}_n$ contains each $b \in B_I$ with a prespecified probability. Note also that there has been much recent interesting work on inference in conditional moment inequality models that might also be adapted to fit this setup. See for example Andrews and Shi (2007), Chernozhukov, Lee, and Rosen (2009), Kim (2007), and Ponomareva (2010).

To illustrate the procedure, assume that x_i has discrete support $\mathcal{X}$ and let $x = (x_1', x_2')'$ be any point in the support where we assume that $P(x_i = x)$ is bounded away from zero and one. Suppose we want to test whether a given β belongs to the identified set and take as an example **Model 1**. Model 2 tests can be done similarly.
According to Theorem 3.1, for a candidate value b to belong to the identified set B_I, the following two inequalities must be satisfied for all τ and x:

$$\begin{aligned} P\{y_{i1}^U - x_{i1}'b \le \tau | x_i\} &\le P\{y_{i2}^L - x_{i2}'b \le \tau | x_i\} \\ P\{y_{i2}^U - x_{i2}'b \le \tau | x_i\} &\le P\{y_{i1}^L - x_{i1}'b \le \tau | x_i\} \end{aligned} \tag{5.1}$$

Or in other words, one random variable conditionally (first order) stochastically dominates another. Again, here, we follow the interesting work of (Linton, Song, and Whang (2010)) and define a test statistic that is based on the KS metric. Other test statistics are possible.

First, define empirical analogs of the above conditional probabilities:

$$\hat{P}^0_{1,n}(\tau,x) = \frac{\sum_{i=1}^n 1\{y^0_{i1} - x'_{i1}b \le \tau\}1\{x_i = x\}}{\sum_{i=1}^n 1\{x_i = x\}}$$
$$\hat{P}^1_{1,n}(\tau,x) = \frac{\sum_{i=1}^n 1\{y^1_{i1} - x'_{i1}b \le \tau\}1\{x_i = x\}}{\sum_{i=1}^n 1\{x_i = x\}}$$
$$\hat{P}^0_{2,n}(\tau,x) = \frac{\sum_{i=1}^n 1\{y^0_{i2} - x'_{i1}b \le \tau\}1\{x_i = x\}}{\sum_{i=1}^n 1\{x_i = x\}}$$
$$\hat{P}^1_{2,n}(\tau,x) = \frac{\sum_{i=1}^n 1\{y^1_{i2} - x'_{i1}b \le \tau\}1\{x_i = x\}}{\sum_{i=1}^n 1\{x_i = x\}}$$

As in models in moment inequalities, the asymptotic distribution of test statistics are non-degenerate only on the "boundary" which we call here "contact sets" as in Linton, Song, and Whang (2010). These are the sets (as a function of x) where the inequalities above bind and are defined as:

$$B^{12}(x) = \left\{\tau : |P^1_{1,n}(\tau,x) - P^0_{2,n}(\tau,x)| = 0\right\}$$
$$B^{21}(x) = \left\{\tau : |P^1_{2,n}(\tau,x) - P^0_{1,n}(\tau,x)| = 0\right\}$$

Finally, define the following test statistic:

$$T_n(b) = \sqrt{n}\sum_{x\in\mathcal{X}}\left[\max\left\{\sup_\tau(\hat{P}^1_{1,n}(\tau,x) - \hat{P}^0_{2,n}(\tau,x)),0\right\} + \max\left\{\sup_\tau(\hat{P}^1_{2,n}(\tau,x) - \hat{P}^0_{1,n}(\tau,x)),0\right\}\right]$$

It is easy to find the asymptotic distribution of the above test statistic since it is a continuous function of sample mean like functions. Let

$$v_{12,n}(\tau,x) = \sqrt{n}\left\{(\hat{P}^1_{1,n}(\tau,x) - \hat{P}^0_{2,n}(\tau,x)) - (P^1_{1,n}(\tau,x) - P^0_{2,n}(\tau,x))\right\}$$
$$v_{21,n}(\tau,x) = \sqrt{n}\left\{(\hat{P}^1_{2,n}(\tau,x) - \hat{P}^0_{1,n}(\tau,x)) - (P^1_{2,n}(\tau,x) - P^0_{1,n}(\tau,x))\right\}$$

Then under standard conditions, $v_{12,n}(\tau,x)$ and $v_{21,n}(\tau,x)$ converge uniformly on $T\times\mathcal{X}$ to gaussian processes $v_{12}(\tau,x)$ and $v_{21}(\tau,x)$ with continuous sample paths. Under the null hypothesis it is easy to show that

$$T_n \xrightarrow{d} \begin{cases} \sum_{x\in\mathcal{X}}\left[\max\left\{\sup_{\tau\in B^{12}(x)} v_{12}(\tau,x),0\right\} + \max\left\{\sup_{\tau\in B^{21}(x)} v_{21}(\tau,x),0\right\}\right] & \text{if } B^{12}(x) \text{ or } B^{21}(x) \text{ is nonempty} \\ 0 \text{ if both } B^{12}(x) \text{ and } B^{21}(x) \text{ are empty} \end{cases}$$

This asymptotic distribution is non-degenerate on the contact sets. These sets need to be estimated, and a possible estimate is:

$$\hat{B}_n^{12}(x) = \left\{\tau : |\hat{P}_{1,n}^1(\tau,x) - \hat{P}_{2,n}^0(\tau,x)| < c_n\right\}$$
$$\hat{B}_n^{21}(x) = \left\{\tau : |\hat{P}_{2,n}^1(\tau,x) - \hat{P}_{1,n}^0(\tau,x)| < c_n\right\}$$

where $c_n \to 0$ and $\sqrt{n}c_n \to \infty$ (e.g. $c_n = \sqrt{\log(n)}$).

Now, define the confidence region for β as the set B_N:

$$\mathcal{B}_n = \{b \in B : T_n(b) \leq d_{(1-\alpha)}\} \tag{5.2}$$

where $d_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of non-degenerate limit of $T_n(b)$. The problem with the above asymptotic distribution is that it is not easy to simulate and so $d_(1-\alpha)$ is not easy to estimate, and so we, as in (Linton, Song, and Whang (2010)), use the bootstrap to approximate d. We describe next the bootstrap procedure approximates the above distribution consistently.

Bootstrap Procedure:

1. Let $w_i = (v_{i1}, v_{i2}, d_{i1}, d_{i2}, x_i)$. Draw $\{\{w_i^*, i = 1, \ldots, n\}\}_{r=1}^R$ from $\{w_i, i = 1, \ldots, n\}$ randomly with replacement. Then, for every draw r, construct Bootstrap versions of $v_{12}(\tau,x)$ and $v_{21}(\tau,x)$ as follows:

$$\hat{P}_{1,n}^{*0}(\tau,x;b) = \frac{\sum_{i=1}^n 1\{y_{i1}^{*0} - x_{i1}^{*'}b \leq \tau\}1\{x_i^* = x\}}{\sum_{i=1}^n 1\{x_i^* = x\}}$$
$$\hat{P}_{1,n}^{*1}(\tau,x;b) = \frac{\sum_{i=1}^n 1\{y_{i1}^{*1} - x_{i1}^{*'}b \leq \tau\}1\{x_i^* = x\}}{\sum_{i=1}^n 1\{x_i^* = x\}}$$
$$\hat{P}_{2,n}^{*0}(\tau,x;b) = \frac{\sum_{i=1}^n 1\{y_{i2}^{*0} - x_{i1}^{*'}b \leq \tau\}1\{x_i^* = x\}}{\sum_{i=1}^n 1\{x_i^* = x\}}$$
$$\hat{P}_{2,n}^{*1}(\tau,x;b) = \frac{\sum_{i=1}^n 1\{y_{i2}^{*1} - x_{i1}^{*'}b \leq \tau\}1\{x_i^* = x\}}{\sum_{i=1}^n 1\{x_i^* = x\}}$$

and we to re-center the process:

$$v_{12,n}^*(\tau,x) = \sqrt{n}\left\{(\hat{P}_{1,n}^{*1}(\tau,x) - \hat{P}_{2,n}^{*0}(\tau,x)) - (\hat{P}_{1,n}^1(\tau,x) - \hat{P}_{2,n}^0(\tau,x))\right\}$$
$$v_{21}^*(\tau,x) = \sqrt{n}\left\{(\hat{P}_{2,n}^{*1}(\tau,x) - \hat{P}_{1,n}^{*0}(\tau,x)) - (\hat{P}_{2,n}^1(\tau,x) - \hat{P}_{1,n}^0(\tau,x))\right\}$$

2. Finally, let

$$T_n^*(b) = \sum_{x\in\mathcal{X}} \left[\max\left\{\sup_{\tau\in\hat{B}_n^{12}(x)} v_{12,n}^*(\tau,x), 0\right\} + \max\left\{\sup_{\tau\in\hat{B}_n^{21}(x)} v_{21,n}^*(\tau,x), 0\right\}\right]$$

   where $\sup_{\tau\in\hat{B}_n^{12}(x)} v_{12,n}^*(\tau,x)$ is defined to be zero if $\hat{B}_n^{12}(x) = \varnothing$ and similarly $\sup_{\tau\in\hat{B}_n^{21}(x)} v_{21,n}^*(\tau,x)$ is defined to be zero if $\hat{B}_n^{21}(x) = \varnothing$.

3. Repeat above 2 steps R times to obtain the empirical distribution of $T_n^*(b)$.

Then a bootstrap confidence region for β can be defined as

$$\hat{B}_n = \{b \in B : T_n(b) \leq \hat{d}^*_{(1-\alpha)}(b)\} \tag{5.3}$$

where

$$\hat{d}^*_{1-\alpha}(b)$$

is the $(1-\alpha)$ quantile of the empirical distribution of $T_n^*(b)$.

Lemma A3 in (Linton, Song, and Whang (2010)) applies in our case, and the proof of consistency of the bootstrap procedure should be similar to the proof of Theorem 2. So, as with the rest of the literature, to construct a confidence regions for β, we collect all the parameters that cannot be rejected using the test statistic above.

5.2 Inference when Model 1 is Point Identified

Under the conditions in Theorem 3.2, Model 1 point identifies β. Under the conditions of this Theorem, we can show (See Proof of Theorem in Appendix) that:

$$E[d_{i1}^U - d_{i2}^L | x_i, \tau_{i1}, \tau_{i2}] > 0 \text{ if and only if } \Delta\tau_i > \Delta x_i'\beta, \tag{5.4}$$

where again, $\tau = (\tau_1, \tau_2)$ is independent of x and α. This rank condition is useful since it is in terms on observed variables and so a variety of estimators can be employed to estimate β consistently. For example, a maximum score style estimator would maximize the following objective function:

$$Q_n(b) = \frac{1}{n}\sum_{i=1}^{n} I[d_{i1}^U > d_{i2}^L] I[\Delta\tau_i > \Delta x_i' b].$$

This is for just one cut point for each cross sectional unit, but for efficiency and root-n consistency we would want to have several such points per unit. This would result in an objective function that looks like a second order U-process:

$$Q_{2n}(b) = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} I[d_{ij1}^U > d_{ij2}^L] I[\Delta\tau_{ij} > \Delta x_i'\beta],$$

where $d_{ij1}^U = I\{y_{i1}^U \leq \tau_{ij1}\}$ and $d_{ij2}^L = I\{y_{i2}^L \leq \tau_{ij2}\}$.

More generally, one could take a continuum of cut point values, in what would be regarded as an "integrated maximum score" procedure. To illustrate Let τ denote a 2×1 vector, whose components are denoted by τ_1, τ_2; let $\omega(\cdot)$ denote a weighting function, say a probability density function on R^2, that integrates to 1. Noting that $d_{11i}, d_{12i}, d_{02i}, d_{01i}$ each depend on cut points τ_1, τ_2, define the function

$$\mathcal{G}_i(\beta) = \int \omega(\tau) I[d_{1i}^U > d_{2i}^L] I[\Delta x_i'\beta > \Delta\tau] + I[d_{2i}^U > d_{1i}^L] I[\Delta x_i'\beta < \Delta\tau] d\tau$$

Averaging this across cross sectional units results the *integrated maximum score* objective function:

$$Q_{In}(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{G}_i(\beta)$$

$\mathcal{G}_i(\cdot)$ can easily be simulated, especially when $\omega(\cdot)$ corresponds to a density function, in which case draws from this distribution can be simulated, and values of the integrand in the definition of $\mathcal{G}_i$ can be averaged across draws.

Under conditions for point identification, the maximizer of any of the above objective functions will converge at the parametric rate to the singleton value β_0 with a limiting normal distribution, from standard results on M or U statistic estimation theory, such as found in, e.g. Newey and McFadden (1994).

However, if the conditions for point identification are not satisfied, the maximizer of the objective function will converge to a set, but one that is larger than the *sharp* set.

6 Extension: 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), there are very few results for censored models like those considered here. Honoré (1993), Honoré and Hu (2004), and Hu (2002), provided results for panel data dynamics with *fixed censoring*, 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 panel model with the censoring structures considered previously. For these models, dynamic feedback can be allowed for in different ways, and this section considers three important cases. The first two will model lagged *observable* dependent variables, and the third will model a lagged *latent* dependent variable. The analysis in this section is mostly heuristic and meant to indicate that our previous approach to analyzing the identified feature in a censored dynamic model can be extended to dynamic setups.

6.1 Dynamics with lagged observed outcomes under Model 1

The first dynamic panel model we consider is one with a *lagged observed dependent variable* as follows:

$$y_{it}^* = \gamma v_{it-1} + x_{it}'\beta + \alpha_i + \epsilon_{it} \tag{6.1}$$

where, again, $v_{it} = \max(y_{it}^*, c_{it})$ is observed. Here 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}, but the analysis again for the independence case is similar. 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 v_{it-1} is last period's observed wage (regardless whether i was employed in sector 1 or 2).

Recall, to accommodate the general random censoring considered in the previous section, again we assume the econometrician does not generally observe y_{it}^*, but does observe the random variables d_{it}, v_{it}, and x_i.

As we show here, our identification approach used in previous sections, based on bounding latent dependent variables, readily extends to the dynamic censored panel models considered here. To illustrate our approach for identifying β and γ, we will make the "initial conditions" assumption that y_{i0}^* is observed, as was the case, in, e.g. Hu (2002). We will now be conditioning on x_i, α_i, and y_{i0}. For the first two periods we have:

$$y_{i1}^* = \alpha_i + \gamma y_{i0}^* + x_{i1}'\beta + \epsilon_{i1} \tag{6.2}$$

$$y_{i2}^* = \alpha_i + \gamma v_{i1} + x_{i2}'\beta + \epsilon_{i2} \tag{6.3}$$

Then we can bound $y_{it}^* - \gamma v_{i,t-1}$ as

$$y_{i10} - \gamma y_{i0}^* \leq y_{i1}^* - \gamma y_{i0}^* \leq y_{i11} - \gamma y_{i0}^* \tag{6.4}$$

$$y_{i20} - \gamma g v_{i1} \leq y_{i2}^* - \gamma v_{i1} \leq y_{i21} - \gamma v_{i1} \tag{6.5}$$

Note that we used y_{i0}^* in the first set of inequalities above since it is observed (i.e. $v_{i0} = y_{i0}^*$). Now, we can subtract the indexes $x_{i1}'\beta, x_{i2}'\beta$ and construct conditional moment inequalities analogous to the construction we had before. In particular, for a candidate (g, b) the following inequalities must hold for all $\tau \in (\infty, +\infty)$ and all values of $\mathbf{x}_i$ and y_{i0}^* in the support:

$$\begin{aligned} P\{\tilde{y}_{i1}^U(g) - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \leq P\{\tilde{y}_{i2}^L(g) - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \\ P\{\tilde{y}_{i2}^U(g) - x_{i2}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \leq P\{\tilde{y}_{i1}^L(g) - x_{i1}'b \leq \tau | \mathbf{x}_i, y_{i0}^*\} \end{aligned} \tag{6.6}$$

where we define $\tilde{y}_{i1}^U(g) \equiv y_{i11} - g y_{i0}^*$, $\tilde{y}_{i2}^U(g) \equiv y_{i21} - g v_{i1}$ and the other terms $\tilde{y}_{i1}^L(g), \tilde{y}_{i2}^L(g)$, analogously. It is easy to show that, as before, the values of b and g that satisfy the above conditional moment inequalities for all x_i, y_{i0}^* and τ will coincide with the *sharp* set. In the next Section, we simulate a version of the above model and examine the identified sets there. Next, we examine another version of the dynamic model.

6.2 Dynamics with lagged sector indicator under Model 1

The second dynamic panel data model we consider is one with a lagged value of the sector variable d_{it} as an explanatory variable, and we maintain the initial conditions assumption as before. This is an interesting model where dynamics of the outcome process is through the sector specific lagged variable. Specifically, for the first two periods we have:

$$y_{i1}^* = \alpha_i + \gamma + x_{i1}'\beta + \epsilon_{i1} \tag{6.7}$$

$$y_{i2}^* = \alpha_i + \gamma d_{i1} + x_{i2}'\beta + \epsilon_{i2} \tag{6.8}$$

Here we have the following inequalities:

$$y_{i10} - \gamma \leq y_{i1}^* - \gamma \leq y_{i11} - \gamma \tag{6.9}$$

$$y_{i20} - \gamma d_{i1} \leq y_{i2}^* - \gamma d_{i1} \leq y_{i21} - \gamma d_{i1} \tag{6.10}$$

Once again, for a candidate value (g, b), we can subtract the indexes $x_{i1}'b, x_{i2}'b$ and construct conditional moment inequalities analogous to before:

$$\begin{aligned} P\{\tilde{y}_{i1}^U(g) - x_{i1}'b \leq \tau|\mathbf{x}_i\} \leq P\{\tilde{y}_{i2}^L(g) - x_{i2}'b \leq \tau|\mathbf{x}_i\} \\ P\{\tilde{y}_{i2}^U(g) - x_{i2}'b \leq \tau|\mathbf{x}_i\} \leq P\{\tilde{y}_{i1}^L(g) - x_{i1}'b \leq \tau|\mathbf{x}_i\} \end{aligned} \tag{6.11}$$

where we define $\tilde{y}_{i1}^U(g) \equiv y_{i11} - g$, $\tilde{y}_{i2}^U(g) \equiv y_{i21} - gd_{i1}$ and the other terms $\tilde{y}_{i1}^L(g), \tilde{y}_{i2}^L(g)$, analogously. As before, it is easy to show that the values of b and g that satisfy the above conditional moment inequalities for all x_i and τ will coincide with the *sharp* set. This sharp set is simulated for a particular version of the model in the next Section. Finally, we discuss the version of the dynamic model with lagged values of the latent outcome.

6.3 Dynamics with lagged latent outcome under Model 1

The third model our inequality approach can be applied to is when the lagged value of the latent variable y_{it}^* is an explanatory variable. Maintaining the initial conditions assumption for the first two periods we have

$$y_{i1}^* = \alpha_i + \gamma y_{i0}^* + x_{i1}'\beta + \epsilon_{i1} \tag{6.12}$$

$$y_{i2}^* = \alpha_i + \gamma y_{i1}^* + x_{i2}'\beta + \epsilon_{i2} \tag{6.13}$$

Here the inequalities become more complicated than before because we do not necessarily observe the right hand side variables. One approach (assuming γ is nonnegative) is to work with the following inequalities:

$$y_{i10} - \gamma y_{i0}^* \le y_{i1}^* - \gamma y_{i0}^* \le y_{i11} - \gamma y_{i0}^* \tag{6.14}$$

$$y_{i20} - \gamma y_{i11} \le y_{i2}^* - \gamma y_{i1}^* \le y_{i21} - \gamma y_{i10} \tag{6.15}$$

Note that we used y_{i0}^* in the first set on inequalities above since it is observed. But y_{i1}^* is not, so in the second set of the above inequalities, we subtracted y_{i11}, y_{i10}. Now, we can subtract the indexes $x_{i1}'\beta, x_{i2}'\beta$ and construct conditional moment inequalities analogous to before. That is, for any τ and x_i, y_{i0}^* the following inequalities must hold:

$$\begin{aligned} &P\{\tilde{y}_{i1}^U(g) - x_{i1}'b \le \tau | \mathbf{x}_i, y_{i0}^*\} \le P\{\tilde{y}_{i2}^L(g) - x_{i2}'b \le \tau | \mathbf{x}_i, y_{i0}^*\} \\ &P\{\tilde{y}_{i2}^U(g) - x_{i2}'b \le \tau | \mathbf{x}_i, y_{i0}^*\} \le P\{\tilde{y}_{i1}^L(g) - x_{i1}'b \le \tau | \mathbf{x}_i, y_{i0}^*\} \end{aligned} \tag{6.16}$$

where we define $\tilde{y}_{i1}^U(g) \equiv y_{i11} - g y_{i0}^*$, and the other terms $\tilde{y}_{i1}^L(g), \tilde{y}_{i2}^L(g), \tilde{y}_{i2}^U(g)$ analogously.

Interestingly and unfortunately, in this case the values of b and g that satisfy the above conditional moment inequalities for all x_i, y_{i0}^* and τ will not generally coincide with the *sharp* set. That happens because we treat bounds on $y_{i1}^* - g y_{i0}^*$ and $y_{i2}^* - g y_{i1}^*$ as independent, while the bounds on $y_{i2}^* - g y_{i1}^*$ depend on the value of y_{i1}^* within the bounds on $y_{i1}^* - g y_{i0}^*$. The only case when the two bounds actually are independent (and therefore conditional moment inequalities in (6.16) give the sharp set) is when it is known that $\gamma = 0$. In all other cases the set defined by (6.16) is too large. Attaining a sharp set in this model is left for future work.

In the dynamic analog of the nonstationary case (i.e. when ϵ_{it} are independent from α_i and $x_{it}()$, it is still possible to construct a set of conditional inequalities that is sharp when $T = 2$. In particular, we can subtract first period equation from the second period:

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

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

$$y_{i20} - (1+\gamma) y_{i11} + \gamma y_{i0}^* \le y_{i2}^* - (1+\gamma) y_{i1}^* + \gamma y_{i0}^* \le y_{i21} - (1+\gamma) y_{i10} + \gamma y_{i0}^* \tag{6.17}$$

Again, for a candidate (g, b) (assuming that $1 + g > 0$) we can subtract $\triangle x_i'b$ and check whether the following inequalities hold for any τ, x_i and x_j:

$$P\{\triangle y_i^U(g) - \triangle x_i'b \le \tau | x_i\} \le P\{\triangle y_j^L(g) - \triangle x_j'b \le \tau | x_j\} \tag{6.18}$$

where $\triangle y_i^U(g) = y_{i21} - (1+g)y_{i10} + gy_{i0}^*$ and $\triangle y_j^L(g) = y_{j20} - (1+g)y_{j11} + gy_{j0}^*$. The set of parameters that satisfy (6.18) gives the *sharp* set. However, if $T > 2$ we cannot any longer claim the sharpness of the intersection of the individual sets for $t = 2, 3, \ldots$. The reason is precisely as before: by doing so we ignore the dependence between the bounds on $y_{i2}^* - (1+g)y_{i1}^* + gy_{i0}^*$ and $y_{i3}^* - (1+g)y_{i12}^* + gy_{i1}^*$. Attaining a sharp set in dynamic models with lagged latent outcomes is left for future work.

7 Simulation Results

This section provides evidence on the size of the identified sets in some stylized panel models with censoring. These simulations are meant to shed light on the size of the identified set in some examples, without issues of sample uncertainty (done with "infinite" sample size). These 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. For these models, simple sufficient conditions for point identification require strong restrictions on the support of the regressors (infinite support) or the correlation structure of the errors. All the simulations are based on the two period model

$$y_t^* = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \quad t = 1, 2 \tag{7.1}$$

where $\beta_1 = \beta_0 = 1$. For all the models we simulate, we use two regressors both with a discrete distribution with support on $\{-1, 0, 1\}$. We first simulate various versions of the above under **Model 1** and **Model 2**. we start with **Model 1**.

7.1 Simulating Model 1

For this model, we plot the set of parameters (b_1, b_2) that satisfy the inequalities in (3.2). 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 $c \sim \mathcal{N}(0,1) + .5\varepsilon^2$. 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.

7.2 Simulating 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 nonstationarity 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.

7.3 Simulating Dynamic Models

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

$$y_{it}^* = \gamma_0 v_{it-1} + x_{it}'\beta_0 + \alpha_i + \epsilon_{it} \tag{7.2}$$

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 v_{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:

$$y_{it}^* \quad = \quad \alpha_i + \gamma_0 d_{it} + x_{it}'\beta_0 + \epsilon_{it} \tag{7.3}$$

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.

8 Conclusions

This paper considered 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 in both panel and cross section settings, such as Honoré (1992), Honoré, Khan, and Powell (2002), and Honoré and Powell (1994).

In addition, our characterization of the identified sets are constructive in that they can be estimated from the sample. The proposed inference method was based on conditional moment inequalities that was *adaptive* to point identification conditions in the sense that our objective function was minimized at the identified set or point, depending on the features of the data generating process. In the latter case, root n consistency and asymptotic normality was established under conditions that are standard in the literature. We also provide guidance on how one might construct confidence regions based on recent contributions to the theory of stochastic dominance tests - See Linton, Song, and Whang (2010).

The work here opens areas for future research. For one, our proposed weight function for the moment points was left as arbitrary, as we only imposed that it be positive and integrate to 1. Further study on its effects on asymptotic properties, and the existence of an optimal function needs to be conducted. Also, there are many avenues to pursue in the panel data setting, such as the further consideration (attaining sharp sets) of a dynamic model where lagged latent dependent variables enter as regressors, as well as consideration of models with more time periods, to see how that may shrink the size of the identified region.

References

Andrews, D., and X. Shi (2007): "Inference based on Conditional Moment Inequalities," Yale University Working Paper.

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 (2008): "Sharp Identification Regions in Games," Working Paper.

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. (2010): "Functional Differencing," Working Paper.

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 (2009): "Interesection Bounds: Estimation and Inference," working paper, MIT.

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

Graham, B., and J. Powell (2009): "Identification and Estimation of 'Irregular' Correlated Random Coefficient Models," Working Paper.

Hoderlein, S., and H. White (2009): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," Working Paper.

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.

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. (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.

Newey, W., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden. North Holland.

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. (2009): "Set Identification via Quantile Restrictions in Short Panels," working paper UCL.

A Appendix

A.1 Proof of Theorem 3.1

Suppose that $b \in B_I$. We will construct $\tilde{y}^*_{it}$ and $\tilde{c}_{it}$ such that *(i)* $\tilde{v}_{it} = \max\{\tilde{y}^*_{it}, \tilde{c}_{it}\}$ has the same distribution conditional on x_i as v_{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 x_i. For the ease of presentation, we define $\eta_{it} \equiv \alpha_i + \epsilon_{it}$ and $\tilde{\eta}_{it} \equiv \tilde{\alpha}_i + \tilde{\epsilon}_{it}$.

Note that

$$P\{y^L_{it} - x'_{it}b \leq \tau | x_i\} = P\{\eta_{it} \leq \tau + x_{it}(b - \beta), y^*_{it} > c_{it} | x_i\} + P\{y^*_{it} \leq c_{it} | x_i\}$$

and

$$P\{y^U_{it} - x'_{it}b \leq \tau | x_i\} = P\{\eta_{it} \leq \tau + x_{it}(b - \beta), y^*_{it} > c_{it} | x_i\} + P\{c_{it} - x'_{it}b \leq \tau, y^*_{it} \leq c_{it} | x_i\}$$

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

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

In this case, $\tilde{v}_{it} = v_{it}$ for $t = 1, 2$. We want $P\{\tilde{\eta}_{i1} \leq \tau | x_i\} = P\{\tilde{\eta}_{i2} \leq \tau | x_i\}$. For each $t = 1, 2$, the sharp upper bound on $P\{\tilde{\eta}_{it} \leq \tau | x_i\}$ is $P\{\eta_{it} \leq \tau + x_{it}(b - \beta), y^*_{it} > c_{it} | x_i\} + P\{y^*_{it} \leq c_{it} | x_i\} = P\{y^L_{it} - x'_{it}b \leq \tau | x_i\}$, while the sharp lower bound (over all possible distributions of u_{it} such that $u_{it} \leq c_{it} - x'_{it}b$) is $P\{\eta_{it} \leq \tau + x_{it}(b - \beta), y^*_{it} > c_{it} | x_i\} + P\{c_{it} - x'_{it}b \leq \tau, y^*_{it} \leq c_{it} | x_i\} = P\{y^U_{it} - x'_{it}b \leq \tau | x_i\}$. Any distribution between these upper and lower bounds can be generated by some distribution of u_{it}. Finally, since b satisfies conditional inequalities (3.2), then we can find u_{i1} and u_{i2} distributed in such a way that $P\{\tilde{\eta}_{i1} \leq \tau | x_i\} = P\{\tilde{\eta}_{i2} \leq \tau | x_i\}$. Therefore, b is observationally equivalent to β. □

A.2 Proof of Theorem 3.2

Construct the following random variables: $d^L_{i2} = I\{y^L_{i2} \leq \tau_{i2}\}$ and $d^U_{i1} = I\{y^U_{i1} \leq \tau_{i1}\}$, where τ_{i1} and τ_{i2} satisfy the above condition. Then $E[d^L_{i2} | x_i, \tau_{i2}] = P\{y^L_{i2} \leq \tau_{i2} | x_i, \tau_{i2}\} =$

$1 - P\{y^L_{i2} > \tau_{i2}|x_i, \tau_{i2}\} = 1 - P\{x'_{i2}\beta + \epsilon_{i2} > c_{i2}, x'_{i2}\beta + \epsilon_{i2} > \tau_{i2}|x_i, \tau_{i2}\} = 1 - P\{x'_{i2}\beta + \epsilon_{i2} > \max\{c_{i2}, \tau_{i2}\}|x_{i2}, \tau_{i2}\} = P\{\epsilon_{i2} < \tau_{i2} - x'_{i2}\beta|x_i, \tau_{i2}\}$ Here the last equality follows from the sufficient condition above.

Similarly, $E[d^U_{i1}|x_i, \tau_{i1}] = P\{\max\{x'_{i1}\beta + \epsilon_{i1}, c_{i1} \leq \tau_{i1}\}|x_i, \tau_{i1}\} = P\{x'_{i1}\beta + \epsilon_{i1} \leq \tau_{i1}, c_{i1} \leq \tau_{i1}|x_i, \tau_{i1}\} = P\{\epsilon_{i1} \leq \tau_{i1} - x'_{i1}\beta|x_i, \tau_{i1}\}$.

Finally, taking into account that $\epsilon_{i1} = \varepsilon_{i1} + \alpha_i$ and $\epsilon_{i2} = \varepsilon_{i2} + \alpha_i$ are identically distributed conditional on x_i, we have: $E[d^L_{i2}|x_i, \tau_{i2}] = F(\tau_{i2} - x'_{i2}\beta|x_i)$ and $E[d^U_{i1}|x_i, \tau_{i1}] = F(\tau_{i1} - x'_{i1}\beta|x_i)$, where $F(\cdot|x_i)$ is a c.d.f. of ϵ_{it} conditional on x_i. Now, taking into account that F is a strictly monotone function, we have

$$E[d^U_{i1} - d^L_{i2}|x_i, \tau_{i1}, \tau_{i2}] > 0 \text{ if and only if } \Delta\tau_i > \Delta x'_i\beta, \tag{A.1}$$

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

A.3 Proof of Theorem 4.1

We can re-write lower bound as $LB(\tau, x_i, b) = P\{y^U_{i2} - y^L_{i1} - \triangle x'_i b \leq \tau|x_i\} = P\{y^U_{i2} - y^L_{i1} - \triangle x'_i b \leq \tau, y^*_{i2} > c_{i2}, y^*_{i1} > c_{i1}\} + P\{y^1_{i2} - y^L_{i1} - \triangle x'_i b \leq \tau, y^*_{i2} > c_{i2}, y^*_{i1} < c_{i1}\} + P\{y^U_{i2} - y^L_{i1} - \triangle x'_i b \leq \tau, y^*_{i2} < c_{i2}, y^*_{i1} > c_{i1}\} + P\{y^U_{i2} - y^L_{i1} - \triangle x'_i b \leq \tau, y^*_{i2} < c_{i2}, y^*_{i1} < c_{i1}\} = P\{\triangle\varepsilon_i + \triangle x'_i\beta \leq \tau + \triangle x'_i b, y^*_{i2} > c_{i2}, y^*_{i1} > c_{i1}\} + 0 + P\{c_{i2} - y^*_{i1} \leq \tau + \triangle x'_i b, y^*_{i2} < c_{i2}, y^*_{i1} > c_{i1}\} + 0$. So that

$$\begin{aligned} LB(\tau, x_i, b) = \quad & P\{\triangle\varepsilon_i + \triangle x'_i\beta \leq \tau + \triangle x'_i b, y^*_{i2} > c_{i2}, y^*_{i1} > c_{i1}\} \\ & + P\{c_{i2} - y^*_{i1} \leq \tau + \triangle x'_i b, y^*_{i2} < c_{i2}, y^*_{i1} > c_{i1}\} \end{aligned} \tag{A.2}$$

Similarly, we can re-write upper bound as $UB(\tau, x_j, b) = P\{y^L_{j2} - y^U_{j1} - \triangle x'_j b \leq \tau|x_j\} = P\{y^L_{j2} - y^U_{j1} - \triangle x'_j b \leq \tau, y^*_{j2} > c_{j2}, y^*_{j1} > c_{j1}\} + P\{y^L_{j2} - y^U_{j1} - \triangle x'_j b \leq \tau, y^*_{j2} > c_{j2}, y^*_{j1} < c_{j1}\} + P\{y^L_{j2} - y^U_{j1} - \triangle x'_j b \leq \tau, y^*_{j2} < c_{j2}, y^*_{j1} > c_{j1}\} + P\{y^0_{j2} - y^1_{j1} - \triangle x'_j b \leq \tau, y^*_{j2} < c_{j2}, y^*_{j1} < c_{j1}\} = P\{\triangle\varepsilon_j + \triangle x'_j\beta \leq \tau + \triangle x'_j b, y^*_{j2} > c_{j2}, y^*_{j1} > c_{j1}\} + P\{y^*_{j2} - c_{j1} \leq \tau + \triangle x'_j b, y^*_{j1} < c_{j1}, y^*_{j2} > c_{j2}\} + P\{y^*_{j1} > c_{j1}, y^*_{j2} < c_{j2}\} + P\{y^*_{j1} < c_{j1}, y^*_{j2} < c_{j2}\}$. So that

$$\begin{aligned} UB(\tau, x_j, b) = \quad & P\{\triangle\varepsilon_j + \triangle x'_j\beta \leq \tau + \triangle x'_j b, y^*_{j2} > c_{j2}, y^*_{j1} > c_{j1}\} + P\{y^*_{j2} < c_{j2}\} \\ & + P\{y^*_{j2} - c_{j1} \leq \tau + \triangle x'_j b, y^*_{j2} > c_{j2}, y^*_{j1} < c_{j1}\} \end{aligned} \tag{A.3}$$

Suppose that $b \in B_I$, that is

$$LB(\tau, x_i, b) \leq UB(\tau, x_j, b) \text{ for any } \tau,\ x_i,\ x_j.$$

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

- If $y^*_{i2} > c_{i2}$, $y^*_{i1} > c_{i1}$, then $\tilde{\alpha}_i = \alpha_i + x'_{i1}\beta - x'_{i1}b$, and $\triangle\tilde{\varepsilon}_i = \triangle\varepsilon_i + \triangle x'_i\beta - \triangle x'_i b$.
- If $y^*_{i2} > c_{i2}$, $y^*_{i1} < c_{i1}$, then $\tilde{\alpha}_i = y^*_{i2} - \triangle\tilde{\varepsilon}_i - x'_{i2}b$, and $\triangle\tilde{\varepsilon}_i = \gamma_i(\triangle\varepsilon_i + \triangle x'_i\beta) + (1 - \gamma_i)(y^*_{i2} - c_{i1}) - \triangle x'_i b + u_{i1}$, where $0 \leq \gamma_i \leq 1$ and $u_{i1} \geq 0$.
- If $y^*_{i2} < c_{i2}$, $y^*_{i1} > c_{i1}$, then $\tilde{\alpha}_i = \alpha_i + x'_{i1}\beta - x'_{i1}b$, and $\triangle\tilde{\varepsilon}_i = \lambda_i(\triangle\varepsilon_i + \triangle x'_i\beta) + (1 - \lambda_i)(c_{21} - y^*_{i1}) - \triangle x'_i b - u_{i2}$, where $0 \leq \lambda_i \leq 1$ and $u_{i2} \geq 0$.
- If $y^*_{i2} < c_{i2}$, $y^*_{i1} < c_{i1}$, then $\triangle\tilde{\varepsilon}_i = \triangle\varepsilon_i + \triangle x'_i\beta - \triangle x'_i b - u_{i3}$ and $\tilde{\alpha}_i = \min\{c_{i1} - x'_{i1}b, c_{i2} - \triangle\varepsilon_i - \triangle x'_i\beta + \triangle x'_i b + u_{i3}\} - u_{i4}$, where $-\infty < u_{i3} < +\infty$ and $u_4 \geq 0$.

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

Now, $P\{\triangle\tilde{\varepsilon}_i \leq \tau | x_i\} = P\{\triangle\varepsilon_i + \triangle x'_i\beta \leq \tau + \triangle x'_i b, y^*_{i2} > c_{i2}, y^*_{i1} > c_{i1} | x_i\} + P\{\gamma_i(\triangle\varepsilon_i + \triangle x'_i\beta) + (1 - \gamma_i)(y^*_{i2} - c_{i1}) \leq \tau + \triangle x'_i b - u_{i1}, y^*_{i2} > c_{i2}, y^*_{i1} < c_{i1} | x_i\} + P\{\lambda_i(\triangle\varepsilon_i + \triangle x'_i\beta) + (1 - \lambda_i)(c_{i1} - y^*_{i1}) \leq \tau - \triangle x'_i b + u_{i2}, y^*_{i2} < c_{i2}, y^*_{i1} > c_{i1} | x_i\} + P\{\triangle\varepsilon_i + \triangle x'_i\beta \leq \tau + \triangle x'_i b + u_{i3}, y^*_{i2} < c_{i2}, y^*_{i1} < c_{i1} | x_i\}$.

Then lower (sharp) bound on $P\{\triangle\tilde{\varepsilon}_i \leq \tau | x_i\}$ over all possible distributions of $u_{i1}, u_{i2}, u_{i3}, u_{i4}$, λ_i, and γ_i is equal to $LB(\tau, x_i, b)$, and upper (sharp) bound on $P\{\triangle\tilde{\varepsilon}_j \leq \tau | x_j\}$ is equal to $UB(\tau, x_j, b)$. 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 x_i etc) so that for any τ, x_i, and x_j we have $P\{\triangle\tilde{\varepsilon}_i \leq \tau | x_i\} = P\{\triangle\tilde{\varepsilon}_i \leq \tau | x_j\} = F(\tau)$ for some $F(\tau)$ such that $LB(\tau, x_i, b) \leq F(\tau) \leq UB(\tau, x_j, b)$, and this distribution is independent of x_i. □

A.4 Proof of Theorem 4.2.

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

A.5 Proof of Theorem 4.3

Part 1. Suppose that $b \in B$ is such that $b_k \neq \beta_k$. Then assumption A1(i) imply that $\triangle x_i'(b-\beta)$ and $\triangle x_j'(b-\beta)$ are unbounded on the support of x_i. Therefore, for any $0 < \delta < 1$ and any τ we can find such values of x_i and x_j that $F_{\triangle\varepsilon}(\tau + \triangle x_i'(b-\beta)) - F_{\triangle\varepsilon}(\tau + \triangle x_j'(b-\beta)) > \delta$. Let $q < \delta < 1$. Then we have $F_{\triangle\varepsilon}(\tau + \triangle x_i'(b-\beta)) - F_{\triangle\varepsilon}(\tau + \triangle x_j'(b-\beta)) > q$ for some x_i and x_j, which is a contradiction to A2*(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 A1(ii) ensures that there exist some $\gamma_2 < \gamma_1$ such that the sets $\overline{\mathcal{X}}_{\gamma_1} = \{x_{i,-k} :$ such that $x_i'(b-\beta) = x_{i,-k}(b_{-k} - \beta_{-k} > \gamma_1\}$ and $\underline{\mathcal{X}}_{\gamma_2} = \{x_{j,-k} :$ such that $x_j'(b-\beta) = x_{j,-k}(b_{-k} - \beta_{-k} < \gamma_2\}$ are nonempty. Then there exist $\rho > 0$ and $\tilde{\tau}$ such that $H(x_{i,-k}, x_{j,-k}) \equiv F_{\triangle\varepsilon}(\tilde{\tau} + \triangle x_i'(b-\beta)) - F_{\triangle\varepsilon}(\tilde{\tau} + \triangle x_j'(b-\beta)) > \rho$ on $\mathcal{X}_{\gamma_1,\gamma_2} = \overline{\mathcal{X}}_{\gamma_1} \times \underline{\mathcal{X}}_{\gamma_2}$. Hence, the left-hand side of (4.2) is bounded away from zero for $\tau = \tilde{\tau}$ on $\mathcal{X}_{\gamma_1,\gamma_2}$ for any values of $x_{i,k}$ and $x_{j,k}$ in the support. On the other hand, assumption A2(ii) implies that the right-hand side of (4.2) can be made less than any $\rho > 0$ with a proper choice of $x_{i,k}$ and $x_{j,k}$. Therefore, β is identified relative to any $b \neq \beta$, so that $B_I = \{\beta\}$.□

A.6 Proof of Theorem 4.4

Note first that for any 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:

$$P\{w_i^U - \triangle x_i' b \leq \tau | x_i\} \leq F(\tau) \leq P\{w_j^L - \triangle x_j' b \leq \tau | x_j\}\} \tag{A.4}$$

for some cumulative distribution function F. Inequalities (A.4) imply that if $\tau_2 - \triangle x_j' b \geq \tau_1 - \triangle x_i' b$, then $P\{w_i^U \leq \tau_{1i} | x_i\} \leq P\{w_j^L \leq \tau_{2j} | x_j\}$. 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

$$P\{w_i^U \leq \tilde{\tau} + \triangle\tilde{x}_i b|\tilde{x}_i\} > P\{w_j^L \leq \tilde{\tau} + \triangle\tilde{x}_j b|\tilde{x}_j\} \tag{A.5}$$

Let $\tilde{\tau}_{2j} = \tilde{\tau} + \triangle\tilde{x}_j b$ and $\tilde{\tau}_{1i} = \tilde{\tau} + \triangle\tilde{x}_i b$. Then $\tilde{\tau}_{2j} - \triangle\tilde{x}_j b = \tilde{\tau}_{1i} - \triangle\tilde{x}_i b = \tilde{\tau}$ and $P\{\triangle y_{ui} \leq \tilde{\tau}_{1i}|\tilde{x}_i\} > P\{\triangle y_{lj} \leq \tilde{\tau}_{2j}|\tilde{x}_j\}$. By continuity of τ and strict inequality in (4.1), there exist the set U of positive probability measure such that for any $(\tau_{1i}, \tau_{2i}, x_i, x_j) \in U$ we have:

1. $\tau_{2j} - \triangle x_j' b \geq \tau_{1i} - \triangle x_i' b$,
2. $P\{\triangle y_{ui} \leq \tau_{1i}|x_i\} > P\{\triangle y_{lj} \leq \tau_{2j}|x_j\}$,

so that $Q(b) > 0$, which implies that if $b \notin B_1$, then $Q(b) > 0$. Therefore, $B_I = B_Q$.$\square$

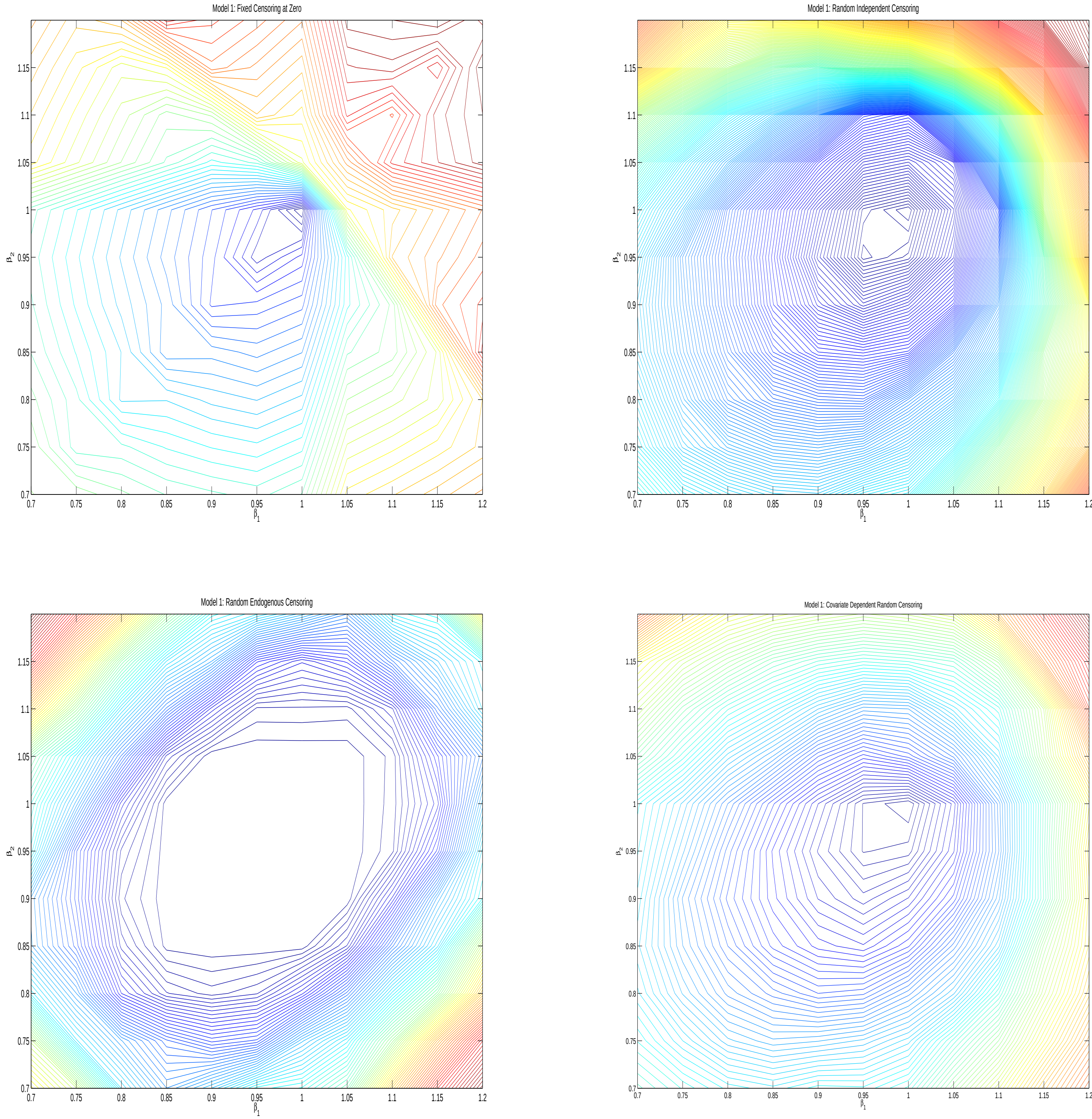

Figure 1: Fixed and Random Independent Censoring (top) Endogenous censoring and covariate dependent censoring (bottom)

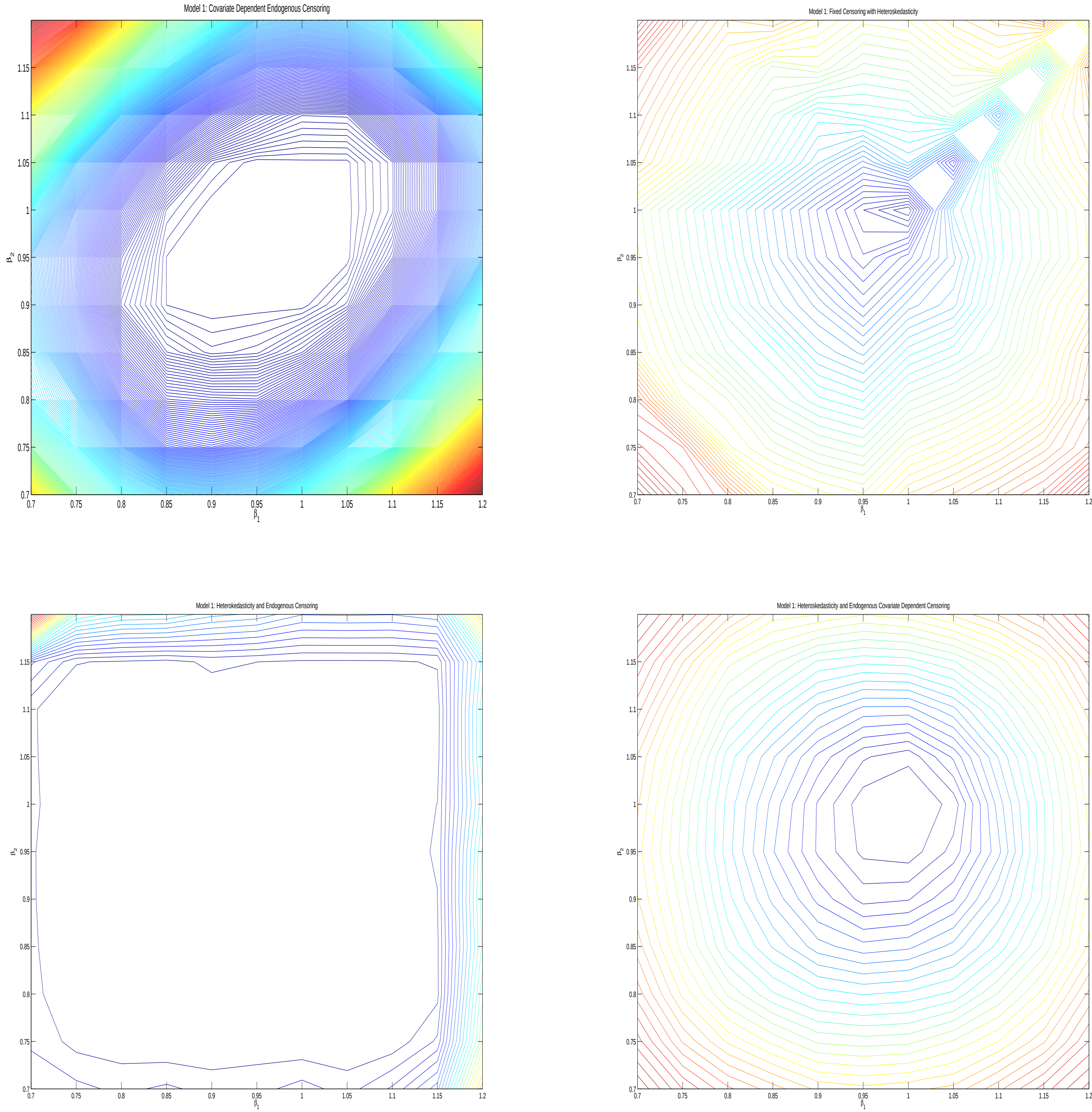

Figure 2: Model 1: Covariate dependent endogenous censoring, fixed censoring with heteroskedasticity (top) Heteroskedastic endogenous censoring and heteroskedastic covariate dependent censoring (bottom)

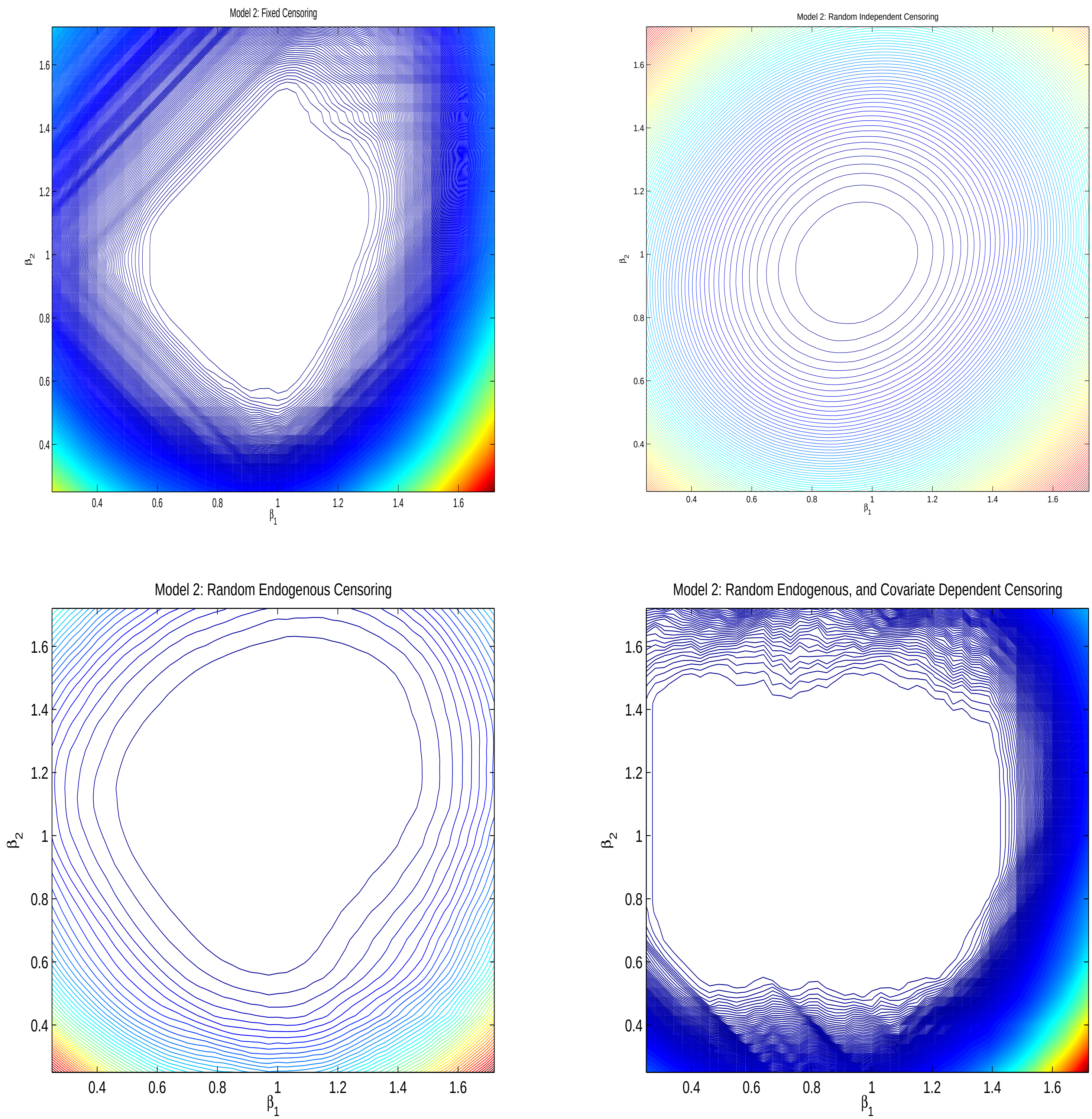

Figure 3: Model 2: Fixed censoring and random independent censoring (top) endogenous censoring and endogenous covariate dependent censoring (bottom)

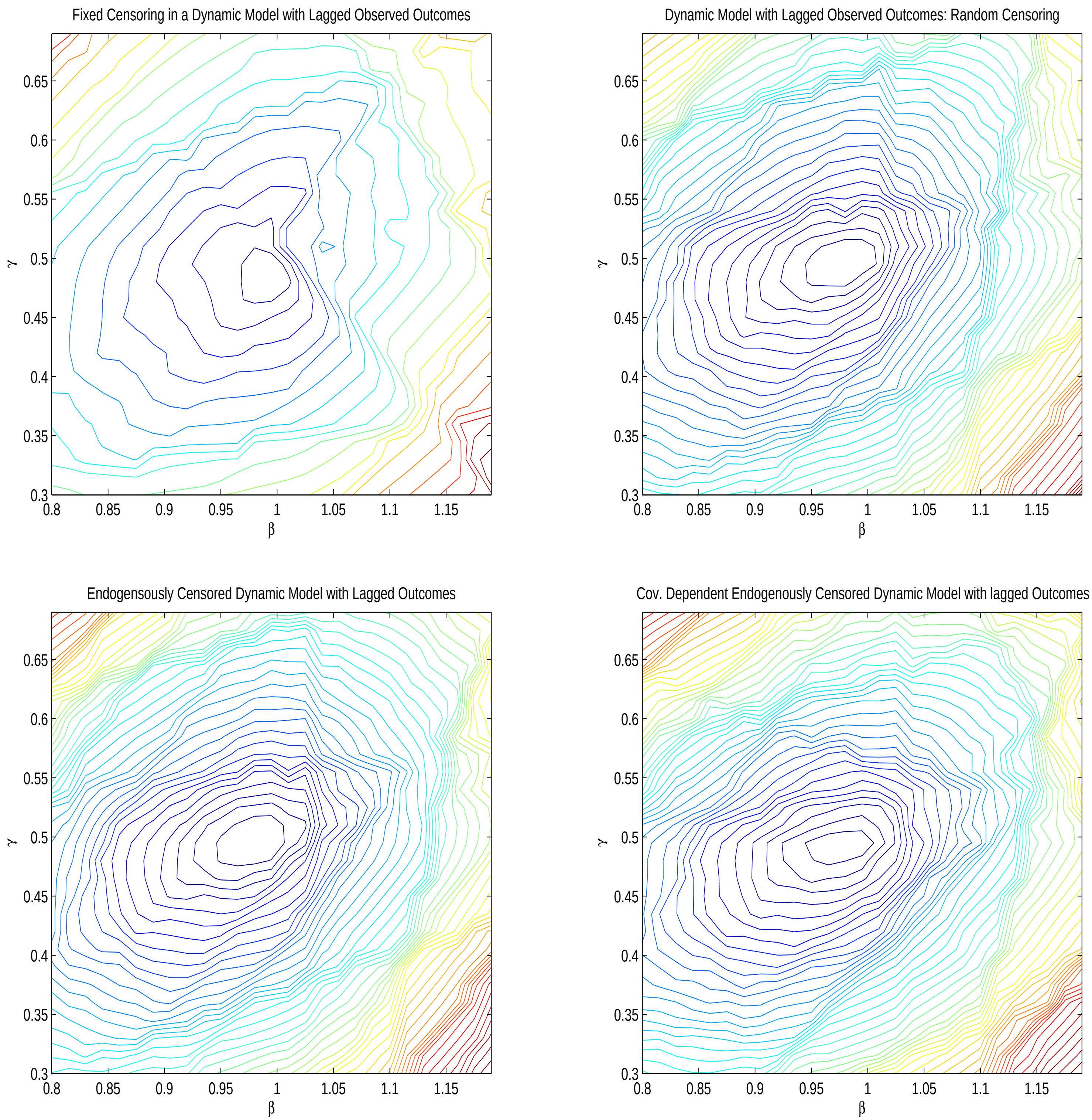

Figure 4: Dynamic Model with Lagged Outcomes: Fixed censoring and random independent censoring (Top) Endogenous and covariate dependent censoring (Bottom)

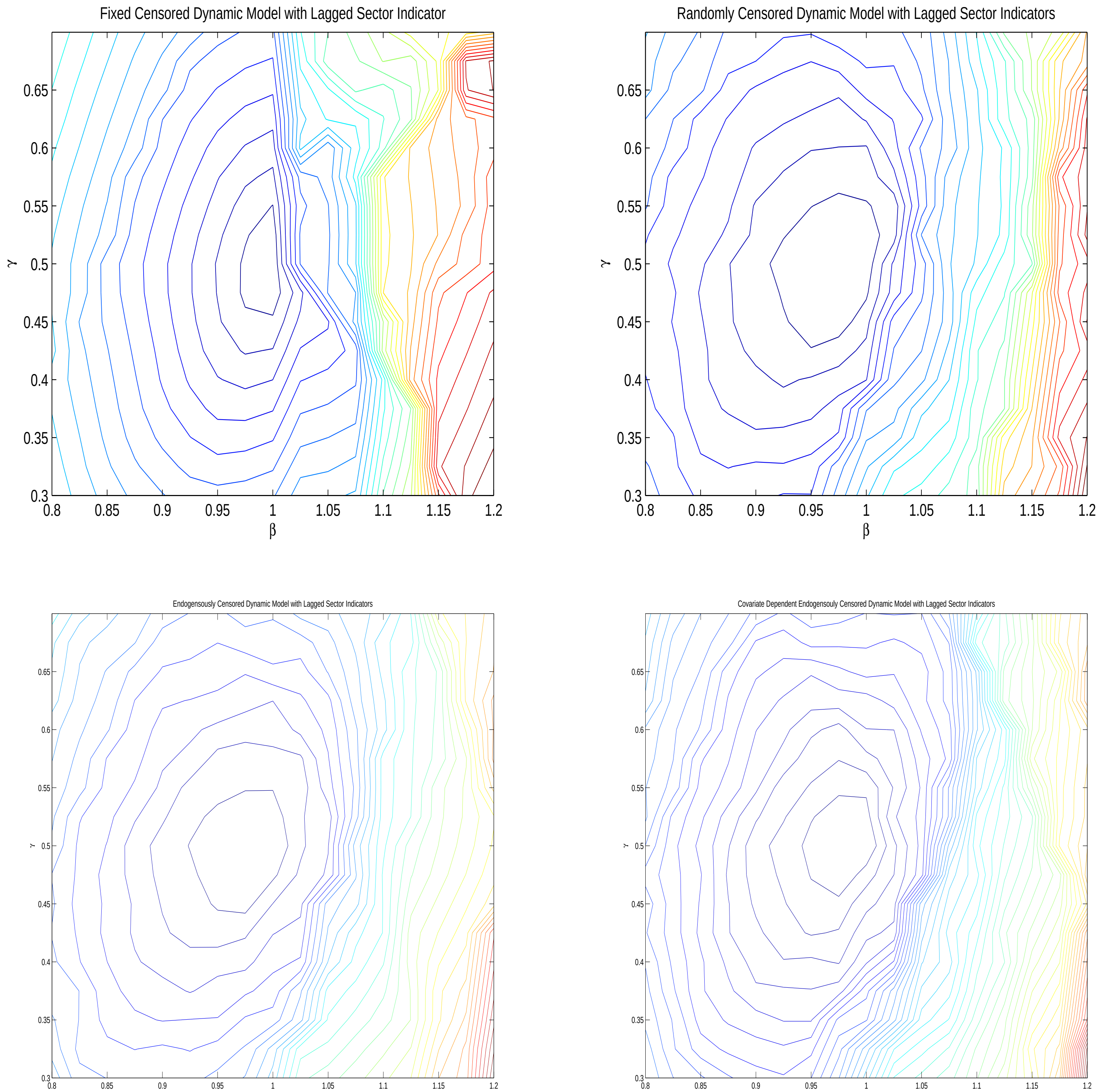

Figure 5: Dynamic Model with Lagged Sector Indicators: Fixed Censoring and Random Independent Censoring (Top) Endogenous Censoring (Bottom)