

Set identification via quantile restrictions in short panels

Adam Rosen

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP26/09

Set Identification via Quantile Restrictions in Short Panels *

Adam M. Rosen[†]
CEMMAP, UCL, and IFS

September 16, 2009

Abstract

This paper studies the identifying power of conditional quantile restrictions in short panels with fixed effects. In contrast to classical fixed effects models with conditional mean restrictions, conditional quantile restrictions are not preserved by taking differences in the regression equation over time. This paper shows however that a conditional quantile restriction, in conjunction with a weak conditional independence restriction, provides bounds on quantiles of differences in time-varying unobservables across periods. These bounds carry observable implications for model parameters which generally result in set identification. The analysis of these bounds includes conditions for point identification of the parameter vector, as well as weaker conditions that result in identification of individual parameter components.

KEY WORDS. Bound analysis, conditional quantiles, partial identification, panel data, fixed effects.

*I thank Ivan Canay, Andrew Chesher, Sokbae Lee, Chuck Manski, Rosa Matzkin, and Elie Tamer for helpful comments and discussion. I have benefited from comments from participants at the May 2009 conference in honor of Chuck Manski's 60th birthday held at Northwestern as well as the CEMMAP conference on quantile regression in London in June 2009. Financial support from the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001 is gratefully acknowledged. The usual disclaimer applies.

[†]email: adam.rosen@ucl.ac.uk

JEL SUBJECT CLASSIFICATION. C14, C21, C23, C50.

1 Introduction

This paper studies the linear panel data model

$$Y_{it} = X_{it}\beta + \alpha_i + U_{it}, \quad (1)$$

with conditional quantile restrictions on U_{it} , where the time dimension is $t = 1, \dots, T < \infty$, and i subscripts observations of individuals in the population. α_i denotes a time-invariant unobserved effect for the i^{th} individual, and U_{it} a time-varying unobservable. The analysis is of the “fixed effect” variety in the sense that the conditional distribution of α_i is left unrestricted. It is well known that the conditional mean restriction $E[U_{it}|x_{i1}, \dots, x_{it}] = 0$ enables application of linear mean regression to the difference of (1) over any two periods for $N^{-1/2}$ -consistent, asymptotically normal estimation of β . Yet when the conditional mean restriction is replaced with a conditional quantile restriction $Q_{U_{it}}(\tau|x_{i1}, \dots, x_{it}) = 0$ for some $\tau \in (0, 1)$, this approach is no longer justified. The non-linearity of the quantile function and in particular its lack of commutativity with subtraction is a substantial complication. As stated by Koenker and Hallock (2000), “Quantiles of convolutions of random variables are rather intractable objects, and preliminary differencing strategies familiar from Gaussian models have sometimes unanticipated effects.” Indeed, in his book on quantile regression, Koenker (2005) places the section on quantile regression with penalized fixed effects in the chapter titled the “Twilight Zone of Quantile Regression.”

This paper provides novel set identification results for the linear fixed effects model with a single conditional quantile restriction $Q_{U_{it}}(\tau|\mathbf{x}_i) = 0$, and finite T , where $\mathbf{x}_i \equiv (x_{i1}, \dots, x_{iT})$. With cross-section data and no fixed effect, conditional quantile restrictions yield point-identification, and corresponding estimators often possess advantages relative to those that employ conditional mean restrictions. In particular, when errors are non-Gaussian, these estimators are often more efficient than those obtained by least squares, see Koenker and Bassett (1978). Yet the analysis of such restrictions with panel data and finite T has so far remained elusive. Again, the complication is that first-differencing does not preserve the quantile restriction in the differenced equation,

as it does with conditional mean restrictions. In this paper it is shown however that conditional quantile restrictions can provide observable implications for model parameters. In particular, I show that in conjunction with a weak conditional independence restriction on time-varying unobservables given \mathbf{x}_i , such restrictions provide bounds on the quantiles of differences in the regression errors across periods. These bounds carry observable implications for model parameters, yielding the derived identification results.

The prior literature on panel data with unobserved effects has not considered conditional quantile restrictions such as $Q_{U_{it}}(\tau|\mathbf{x}_i) = 0$ in isolation, although a number of related models have been studied. First, there are models that do not impose quantile restrictions directly, but whose assumptions imply quantile restrictions on the differenced equation. The semi-parametric binary-response model of Manski (1987) shows that if the idiosyncratic errors U_{it} are stationary conditional on c_i, x_{i1}, x_{i2} , the difference $\text{sgn}(y_{i2}) - \text{sgn}(y_{i1})$ has conditional median $(x_{i2} - x_{i1})\beta$, and the maximum score estimator can be employed for consistent estimation of β . Abrevaya (2000) uses a similar rank-based approach in models with non-binary outcomes. In models with censored or truncated data, Honore (1992) shows how an assumption that errors are conditionally independent and identically distributed implies symmetry of the distribution of observed outcomes, which he uses to identify and consistently estimate the parameters of his model. In the models considered by Geraci and Bottai (2007) and Graham, Hahn, and Powell (2008), where U_{it} are independent of \mathbf{X}_i with the Laplace distribution, the likelihood for ΔY_i given X_i is the negative of the least absolute deviations criterion, thereby justifying application of median regression to first differences in that particular context.

While the approach taken in this paper is similar in nature to that from the semi-parametric panel data literature, e.g. Manski (1987), Honore (1992), and Abrevaya (2000), among others, the assumptions invoked on the unobservables in this paper are comparatively weak. First the marginal distribution of U_{it} , conditional on covariates, need not be identical across t . Indeed, if that were imposed then prior methods could be used to achieve point-identification.¹ Rather, all that is required is that the conditional distribution of U_{it} have *the same* τ -quantile for all t . This is a much weaker restriction than full stationarity, allowing for more flexible patterns of unobserved heterogeneity than just those embodied by the fixed effect. Second, full conditional independence of U_{it} and U_{is} , $s \neq t$, is not required. Instead I employ a weaker restriction on the stochas-

¹If fact, if the distribution of unobservables were integrable, the usual first-differencing approach would apply.

tic relation between time-varying unobservables. Specifically, as laid out in Assumption WCI of Section 2, what is required is that, conditional on covariates, the event that U_{it} falls below its τ -quantile is independent of the event that U_{is} falls below its τ -quantile. This allows for the possibility of some forms of stochastic dependence between U_{it} and U_{is} .

Another important area of related research is the recent literature on quantile regression with panel data, where the goal is to estimate the conditional quantiles of Y_{it} given x_{it} at many quantiles. To date, much of this literature has focused on the pure location-shift model, first considered by Koenker (2004), who provides a penalized quantile regression estimator. The task at hand in this paper is related, but the focus is on identification of the parameters of a single equation, rather than at many different quantiles. Thus, independence is only required for a single quantile. Moreover, the analysis of this paper is for small T , whereas the asymptotic results for the pure location shift model require both N and T going to infinity. In that model, for each τ , the τ quantile of Y_{it} conditional on x_{it} is given by $x_{it}\beta(\tau) + \alpha_i$. The present context is related, although the fixed effect plays a different role. Specifically, in this paper the τ conditional quantile of Y_{it} given x_{it} need not be additively separable in the fixed effect, since the conditional quantile restriction on U_{it} gives $Q_{Y_{it}}(\tau|x_{it}) = x_{it}\beta + Q_{\alpha_i+U_{it}}(\tau|x_{it})$, which is only equivalent if $Q_{\alpha_i+U_{it}}(\tau|x_{it}) = \alpha_i$ for all i, t .

A number of additional papers provide further results in the context of the pure location shift and closely related models. Lamarche (2006) provides additional asymptotic results and an optimal choice for the regularization parameter. Subsequently Lamarche (2008) develops an alternative estimator that allows the fixed effect to be a linear function of \mathbf{x}_i . Canay (2008) shows how a simple transformation can be used to eliminate the fixed effect from the pure location shift model, motivating a simple and easily computed two-step asymptotically normal estimator. Galvao (2008) considers a dynamic panel data model where the fixed effect is a pure location shift, but where lagged dependent variables in the conditional quantile specification cause bias. He shows how the instrumental variable quantile regression estimator of Chernozhukov and Hansen (2006) and Chernozhukov and Hansen (2008) can be used to alleviate the bias. The related paper by Harding and Lamarche (2008) shows how instrumental quantile regression can be employed in panel data models with endogenous regressors, demonstrating favorable performance in a Monte Carlo study and application. Asymptotic analysis in all of

these models, where provided, is conducted with both N and T going to infinity.

Recently, additional approaches to the analysis of quantiles in panel data models have been proposed. For example, Abrevaya and Dahl (2008) employ a correlated random effects approach in the spirit of Chamberlain (1982) in order to estimate conditional quantile functions. The nature of the correlated random effects approach differs fundamentally from that considered here by imposing additional structure on the relationship between the unobserved effect and covariates. Also related are recent papers by Graham and Powell (2008), Graham, Hahn, and Powell (2009), and Arellano and Bonhomme (2009). These papers employ random coefficient approaches that extend the correlated random coefficients framework of Chamberlain (1982) and Chamberlain (1992) to identify and consistently estimate various distributional features of outcomes given covariates. Although these models, and thus the subsequent identification results, are quite different than the model considered here, the restrictions in these papers are also of the fixed effects variety. Other papers that derive bounds on parameters in panel data models under different restrictions than those considered here include Honore and Tamer (2006) and Chernozhukov, Fernandez-Val, Hahn, and Newey (2008).

The rest of the paper proceeds as follows. Section 2 provides the formal model and the identification results. The results are constructive in the sense that the identified set is characterized by a set of inequality restrictions that can provide a basis for estimation and inference. Section 3 briefly discusses how methods from the prior literature on set identification obtained from inequality restrictions can thus be used for consistent estimation of the identified set. Section 4 provides examples, and section 5 concludes. Proofs are collected in the Appendix.

2 The Model and Observable Implications

Let the data consist of observations $\{(y_{it}, x_{it}) : i = 1, \dots, N; t = 1, \dots, T\}$ from the model

$$Y_{it} = X_{it}\beta + \alpha_i + U_{it}, \tag{2}$$

where Y_{it} and U_{it} are random variables for each i, t , and X_{it} is $k \times 1$ random vector. Let $\mathbf{Y}_i \equiv (Y_{it}, \dots, Y_{iT})'$, and $\mathbf{X}_i \equiv (X'_{i0}, \dots, X'_{iT})'$ denote the $T \times 1$ and $T \times k$ matrices of outcomes and covariates for given i across t . Similarly define $\mathbf{U}_i \equiv (U_{it}, \dots, U_{iT})$, as the

vector of time-varying unobservables for individual i over all values of t . Bold font is used throughout to explicitly denote collections of variables across all t . For simplicity it is assumed that the panel is balanced, so that there are observations of (\mathbf{y}, \mathbf{x}) for each of N individuals for each period $t = 1, \dots, T$, but this is easily relaxed. Following standard convention, capital letters are used to denote random variables and lower case letters particular realizations. The following assumptions are employed.

Assumption A1 (random sampling): $\{(\mathbf{Y}_i, \mathbf{X}_i) : i = 1, \dots, N\}$ are iid \mathbb{P} . The parameter β belongs to the compact parameter space B .

Assumption A2 (continuity at 0): $U_{it}|\mathbf{x}_i$ is continuously distributed a.e. \mathbf{X}_i and $f_t(0|\mathbf{x}_i) \geq \epsilon$, for some $\epsilon > 0$, where $f_t(\cdot|\mathbf{x}_i)$ denotes the conditional density of U_{it} given $\mathbf{X}_i = \mathbf{x}_i$.

Assumption CQR (conditional quantile restriction): $Q_{u_{it}}(\tau|\mathbf{x}_i) = 0$ with probability 1 for all $t \in \{0, \dots, T\}$.

Assumption WCI (weak conditional independence): For all $(s, t) \in \{0, \dots, T\}^2$ such that $s \neq t$, the events $\{U_{is} \leq 0\}$ and $\{U_{it} \leq 0\}$ are conditionally independent given x_{is} and x_{it} .

Assumption A1 is a standard assumption of random sampling in the cross-section dimension of the observations and also requires compactness of the parameter space. Assumption A2 is a regularity condition that guarantees uniqueness of the τ conditional quantile of U_{it} given \mathbf{x}_i . The restriction CQR is the conditional quantile restriction, which replaces the standard conditional mean restriction $\mathbb{E}(U_{it}|\mathbf{x}_i) = 0$ in classical fixed effects models, e.g. Wooldridge (2002). Condition WCI is a restriction on the stochastic relationship between U_{is} and U_{it} conditional on observed covariates. It specifies that the events that each of them exceeds their τ -quantile are conditionally independent. As noted in the introduction, this is a weaker restriction than full conditional independence. In section 2.2 below I provide sharp bounds on β under CQR and WCI. These bounds remain valid under conditional stochastic independence, though sharper bounds may be obtainable. One advantage of condition WCI is that it does not altogether rule out serial dependence. It is also worth repeating that none of these assumptions require U_{is} and U_{it} to be identically distributed conditional on the covariates. Were that the case, identification of β could be shown by consideration of the usual first-differenced mean regression if their distribution were integrable, and otherwise by approaches from the semiparametric panel data literature, e.g. Honore (1992) and Abrevaya (2000). No

further assumptions are made regarding the distribution of the fixed effect conditional on \mathbf{x}_i or U_{it} .

2.1 Identification Without Conditional Independence

Before considering the full identifying power of all of the assumptions, I first consider the identifying power of A1, A2, and the conditional quantile restriction alone. There is the following result.

Theorem 1 *Assume A1, A2, and CQR. Then every $b \in B$ is observationally equivalent to β .*

By itself, i.e. without imposing the conditional independence restriction WCI, the conditional quantile restriction has no identifying power, as every element of the parameter space is observationally equivalent to β . This is in fact the case *for any* finite T . The underlying reason is that the conditional quantile restriction does very little to restrict the distribution of the idiosyncratic shocks U_{it} across t . The location restriction that it does imply is sufficiently weak that it can be effectively counter-balanced by appropriate choice of the fixed effect α_i , whose conditional distribution has been left completely unrestricted. Thus, for any observed distribution of (\mathbf{Y}, \mathbf{X}) and any conjectured parameter value b there exists a conditional distribution of α_i such that, conditional on any realization of \mathbf{X}_i , $Y_{it} - X_{it} - \alpha_i \leq 0$ for all t with probability τ . As the proof shows formally, precisely such a latent conditional distribution for α_i can be constructed by requiring that α_i be sufficiently small with probability τ , and sufficiently large for with probability $1 - \tau$. It follows that for non-vacuous set identification of β , additional restrictions are needed. The conditional independence restriction WCI is considered below. Additional restrictions on the distribution of α_i given x_i , i.e. “random effects” analysis, could of course also have identifying power, but are not pursued here.

2.2 Adding Conditional Independence Restrictions

When Assumption WCI is imposed jointly with the conditional quantile restriction, observable implications can be derived from differences in (2) across t , as done with

classical fixed effects, albeit in a different manner. To see how, restrict attention to the case where $T = 2$ and consider the following observation.

$$U_{i2} \leq 0 \text{ and } U_{i1} \geq 0 \Rightarrow \Delta U_i \leq 0, \quad (3)$$

$$U_{i2} > 0 \text{ and } U_{i1} \leq 0 \Rightarrow \Delta U_i > 0, \quad (4)$$

where $\Delta U_i \equiv U_{i2} - U_{i1}$. By first Assumption WCI and then Assumption CQR we have that

$$P \{U_{i2} \leq 0 \wedge U_{i1} \geq 0 | \mathbf{x}_i\} = P \{U_{i2} \leq 0 | \mathbf{x}_i\} P \{U_{i1} \geq 0 | \mathbf{x}_i\} = \tau (1 - \tau).$$

$$P \{U_{i2} > 0 \wedge U_{i1} \leq 0 | \mathbf{x}_i\} = P \{U_{i2} > 0 | \mathbf{x}_i\} P \{U_{i1} \leq 0 | \mathbf{x}_i\} = \tau (1 - \tau),$$

where \wedge denotes the logical “and” operator. Combining this with (3) and (4), it follows that

$$P \{\Delta U_i \leq 0 | \mathbf{x}_i\} \geq \tau (1 - \tau),$$

$$P \{\Delta U_i > 0 | \mathbf{x}_i\} > \tau (1 - \tau),$$

or equivalently,

$$\tau (1 - \tau) \leq \Pr \{\Delta U_i \leq 0 | \mathbf{x}_i\} \leq 1 - \tau (1 - \tau). \quad (5)$$

The quantity ΔU_i is a function of observable quantities and the parameter β from which the fixed effect α_i is absent. (5) thus provides bounds on an identified function of β , namely $P \{\Delta U_i \leq 0 | \mathbf{x}_i\}$. This quantity may vary with \mathbf{x}_i but must lie between $\tau (1 - \tau)$ and $1 - \tau (1 - \tau)$ for all \mathbf{x}_i , delivering the following theorem.

Theorem 2 *Let Assumptions A1, A2, CQR, and WCI hold, and assume $T = 2$. Define the set*

$$S(b) \equiv \{\mathbf{x}_i : \Delta x_i b < Q_{\Delta Y_i}(\tau(1 - \tau) | \mathbf{x}_i) \vee \Delta x_i b > Q_{\Delta Y_i}(1 - \tau(1 - \tau) | \mathbf{x}_i)\}, \quad (6)$$

where \vee denotes the logical “or” operator. Then b is identified relative to β if and only if $P \{S(\mathbf{x}; b)\} > 0$. Equivalently the sharp identified for β is

$$B_I \equiv \{b \in B : Q_{\Delta Y_i}(\tau(1 - \tau) | \mathbf{x}_i) \leq \Delta x_i b \leq Q_{\Delta Y_i}(1 - \tau(1 - \tau) | \mathbf{x}_i) \text{ a.e. } \mathbf{X}_i\}. \quad (7)$$

Corollary 1 B_I is non-empty and convex.

Theorem 2 provides observable implications that can be used to distinguish some values of $b \in B$ from the true parameter value β . Equivalently, it provides bounds on the parameter vector β , namely the convex set B_I . This set is sharp. That is, for every $b \in B_I$, and any value of the covariates \mathbf{x}_i , there exists a conditional distribution of the unobservables given \mathbf{x}_i , such that (2) with $\beta = b$ generates the observed distribution of (\mathbf{Y}, \mathbf{X}) .

When $T > 2$, it is straightforward to generalize these bounds by considering the set of parameter values that satisfy the derived inequality restrictions across all time period pairs $s \neq t$, $\{s, t\} \in \{1, \dots, T\}^2$. An immediate corollary of Theorem 2, provided below, is that the intersection of the bounds B_I taken across all such pairs provides bounds on β . For the statement of the result, define $\Delta_{ts}Y_i \equiv Y_{it} - Y_{is}$, $\Delta_{ts}X_i \equiv X_{it} - X_{is}$, and $\Delta_{ts}U_i \equiv U_{it} - U_{is}$.

Corollary 2 If $T > 2$, then $\beta \in B_I$, where

$$B_I \equiv \bigcap_{1 \leq s < t \leq T} B_I^{st},$$

and

$$B_I^{st} \equiv \{b \in B : Q_{\Delta_{st}Y_i}(\tau(1-\tau) | x_{is}, x_{it}) \leq \Delta_{ts}x_i b \leq Q_{\Delta_{st}Y_i}(1-\tau(1-\tau) | x_{is}, x_{it}) \text{ a.e. } X_{is}, X_{it}\}. \quad (8)$$

Returning to the case where $T = 2$ for ease of notation, it is useful to consider what factors contribute to the size of the set B_I (or each of the sets B_I^{st} when $T > 2$). To this end, it is useful to re-write the inequality restrictions that define B_I as

$$Q_{\Delta U_i}(\tau(1-\tau) | \mathbf{x}_i) \leq \Delta x_i (b - \beta) \leq Q_{\Delta U_i}(1-\tau(1-\tau) | \mathbf{x}_i) \text{ a.e. } \mathbf{X}_i, \quad (9)$$

which follows from substituting $\Delta x_i \beta + \Delta U_i$ for ΔY_i . From this it can first be seen that the closer are the $\tau(1-\tau)$ and $1-\tau(1-\tau)$ conditional quantiles of ΔU_i given \mathbf{x}_i , the more tightly concentrated is the set B_I around β . Since the τ conditional quantile of U_{it} is zero, the lower bounding quantile cannot exceed zero, and the upper bounding quantile can be no less than zero. All else equal, the closer is either bounding quantile

to zero for any \mathbf{x}_i the tighter is the implied bound on β . A second factor that affects the size of B_I is the magnitude of Δx_i . For any constant $c > 1$, if the bounding conditional quantiles were equal when conditioned on either \mathbf{x}_i or $c\mathbf{x}_i$, then fewer values of b satisfy the inequality restrictions at $\mathbf{X}_i = c \cdot \mathbf{x}_i$ than at $\mathbf{X}_i = \mathbf{x}_i$.

These observations motivate the following theorem, which provides sufficient conditions for point-identification. The theorem makes use of the following large support condition on the k -th component of ΔX_i , used previously in a number of papers to establish identification in semiparametric models, for example Manski (1985) and Han (1987).

Assumption C1 (support): The distribution of the k^{th} component of ΔX_i , denoted $\Delta X_{i,k}$ conditional on any realization of all other components, denoted $\Delta x_{i,-k}$, is absolutely continuous on \mathbb{R} with respect to Lebesgue measure and $\beta_k \neq 0$.

Theorem 3 *Let Assumptions A1, A2, CQR, WCI, and C1 hold, and let $T = 2$. Suppose that (i) for any set $\mathcal{X}_{i,-k}$ on the support of $\mathbf{X}_{i,-k}$,*

$$\lim_{c \rightarrow \infty} Q_{\Delta U_i}(\tau(1-\tau) | \Delta x_{i,k} > c, \mathbf{X}_{i,-k} \in \mathcal{X}_{i,-k}) = 0, \quad (10)$$

and

$$\lim_{c \rightarrow \infty} Q_{\Delta U_i}(1-\tau(1-\tau) | \Delta x_{i,k} > c, \mathbf{X}_{i,-k} \in \mathcal{X}_{i,-k}) = 0. \quad (11)$$

and (ii) that the support of ΔX_i is not contained in any proper linear subspace of \mathbb{R}^k . Then $B_I = \{\beta\}$.

The above theorem relies on an identification at infinity argument. Condition (i) states that for any fixed $\mathbf{x}_{i,-k}$, as $\Delta x_{i,k}$ is made arbitrarily large, the bounding conditional quantiles of ΔU_i approach zero. Thus, the inequality conditions (9) essentially provide a moment condition for $\Delta x(b - \beta)$ in the limit as $\Delta x_{i,k} \rightarrow \infty$. Condition (ii) then provides a rank condition that guarantees sufficient variation in covariates to achieve point identification. These conditions may be feasible in some contexts, though they are admittedly strong. Nonetheless, the theorem serves the purpose of illustrating when the identified set shrinks to a single point, thereby helping to illustrate the nature of these bounds.

2.3 Point-Identification of Individual Components

This section considers conditions more widely-applicable than those of Theorem 3 under which a particular component of the parameter vector, β_k , is point-identified. The required conditions are the support condition C1 already stated, and condition C2, provided below. For ease of exposition the conditions and subsequent theorem are again provided for the case $T = 2$, but are easily generalized to cover arbitrary T .

Assumption C2 (thin tails): There exist constants C_0 and C_1 such that for almost every \mathbf{x}_i on the support of \mathbf{X}_i , $Q_{\Delta U_i}(\tau(1-\tau) | \mathbf{x}_i) \geq C_0 > -\infty$ and $Q_{\Delta U_i}(1-\tau(1-\tau) | \mathbf{x}_i) \leq C_1 < \infty$.

Assumption C2 bounds the tails of the conditional distribution of ΔU_i given \mathbf{x}_i . The condition requires that there exists some finite values C_0 and C_1 such that for all \mathbf{x}_i , the $\tau(1-\tau)$ conditional quantile of ΔU_i is no lower than C_0 and its $1-\tau(1-\tau)$ conditional quantile is no greater than C_1 . Note that the magnitudes of C_0 and C_1 may be arbitrarily large. The assumption guarantees that one can not select a sequence of \mathbf{x}_i for which either tail of the conditional distribution of ΔU_i escapes to infinity.

With the addition of conditions C1 and C2 to those of Theorem 2, β_k is point-identified, as formalized by the following theorem.

Theorem 4 *Let A1, A2, CQR, WCI, C1 and C2 hold. Let $b \in B$ with $b_k \neq \beta_k$. Then β is identified relative to b , so that β_k is point-identified.*

The intuition behind this theorem is as follows. Suppose b and the true parameter β differ in their k^{th} component. Then, under the support condition C1 on ΔX_{ik} , it follows that conditional on any value for all covariates excluding k , there exists a positive measure set of values on which Δx_{ik} is sufficiently large that $\Delta x_i(b - \beta)$ lies outside the interval $[C_0, C_1]$. Under assumption C2 this implies a violation of the inequality restrictions that define the identified set given in Theorem 2.

3 Estimation and Inference

As shown in Theorem 2, the identified set B_I is by definition the set of b such that $\Delta x b$ satisfies a pair of inequality restrictions for almost every \mathbf{X}_i . Estimation of sets defined by inequality restrictions is well-studied in the recent econometrics literature

on set identification, and results from papers such as Manski and Tamer (2002) and Chernozhukov, Hong, and Tamer (2007) (henceforth CHT) apply here. In this section, I show how a criterion-based estimator analogous to one previously used for set estimation with conditional moment inequalities can be used to consistently estimate the identified set B_I . I then briefly discuss possible approaches for inference.

Specifically, for each pair $s \neq t$, $\{s, t\} \in \{1, \dots, T\}^2$, let $\hat{Q}_{st}^L(x_{is}, \mathbf{x}_{it})$ and $\hat{Q}_{st}^U(x_{is}, \mathbf{x}_{it})$ be consistent nonparametric estimators for the $\tau(1 - \tau)$ and $1 - \tau(1 - \tau)$ conditional quantiles of $\Delta_{st}Y$ given (x_{is}, x_{it}) ,

$$\hat{Q}_{st}^L(x_{is}, x_{it}) \xrightarrow{P} Q_{\Delta_{st}Y_i}(\tau(1 - \tau) | \mathbf{x}_{is}, x_{it}), \quad (12)$$

and

$$\hat{Q}_{st}^U(x_{is}, x_{it}) \xrightarrow{P} Q_{\Delta_{st}Y_i}(1 - \tau(1 - \tau) | x_{is}, x_{it}). \quad (13)$$

Nonparametric conditional quantile estimation can be implemented by local polynomial approximation as in Chaudhuri (1991) or with smoothing splines as in Koenker, Ng, and Portnoy (1994), for example. Define the loss function

$$G_N(b) \equiv \frac{1}{N} \sum_{1 \leq s < t \leq T} \sum_{i=1}^N \left\{ \left\| \hat{Q}_{st}^L(x_{is}, x_{it}) - \Delta_{ts}x_i b \right\|_+^2 + \left\| \hat{Q}_{st}^U(x_{is}, x_{it}) - \Delta_{ts}x_i b \right\|_-^2 \right\},$$

where $\|z\|_+^2 \equiv z^2 1[z > 0]$ and $\|z\|_-^2 \equiv z^2 1[z < 0]$. This function falls within the class of objective functions considered by Manski and Tamer (2002) and CHT, among others, when identified sets are defined by a set of moment inequalities. The quantity $G_N(b)$ measures the degree to which any value b violates the inequality restrictions defining B_I when the population distribution of \mathbf{Y} given \mathbf{X} is replaced with the observed empirical distribution. It is straightforward to show that $G_N(b)$ converges uniformly in probability to a population criterion function $G(b)$ which attains its minimum only on values of $b \in B_I$. From application of Theorem 3.1 of CHT it then follows that the identified set B_I is consistently estimated by the set estimate

$$\hat{B}_I \equiv \{b \in B : G_N(b) \leq c_n\},$$

where $\sup_{b \in B_I} G_N(b) = O_p(1/a_n)$ and c_n is a sequence of positive constants converging to infinity such that $c_n/a_n \rightarrow 0$, e.g. $c_n = \log n$, when a_n is polynomial in n . In some

cases it may also be feasible to show that CHT's degeneracy property holds, for example when there is point identification, in which case one can select $c_n = 0$.

A variety of approaches may also be used for inference, although the validity of each method will depend on the precise context. When covariates are discrete, the number of inequality restrictions that define the identified set are finite, and potentially applicable methods include those of Andrews and Soares (2007), Beresteanu and Molinari (2008), Bugni (2007), Canay (2007), CHT, Galichon and Henry (2009), Romano and Shaikh (2008), and Rosen (2008). When the covariates are continuous, the restrictions defining the identified set embody uncountably many conditional inequality restrictions. Inference that incorporates infinitely many conditional inequalities is an ongoing area of research. Recent work focuses on conditional moment inequalities, e.g. Andrews and Shi (2009), Chernozhukov, Lee, and Rosen (2009) and Menzel (2009), and under suitable conditions variants of these methods are likely to apply in the present context. In what follows, I consider the applicability of the method of Chernozhukov, Lee, and Rosen (2009), henceforth CLR.

To illustrate the possibility for application of the inferential method of CLR, note that the identified set B_I can be re-written as the set of b that satisfy

$$\sup_{\mathbf{x}_i} \{Q_{st}^L(x_{is}, x_{it}) - \Delta_{st} \mathbf{x}_i b\} \leq 0 \leq \inf_x \{Q_{ts}^U(x_{is}, x_{it}) - \Delta_{ts} \mathbf{x}_i b\},$$

where the supremum and infimum are taken over the support of \mathbf{X}_i . Let

$$\theta_u(x, b) \equiv Q_{st}^U(x_{is}, \mathbf{x}_{it}) - \Delta_{ts} \mathbf{x}_i b,$$

and

$$\theta_l(x, b) \equiv Q_{st}^L(x_{is}, \mathbf{x}_{it}) - \Delta_{ts} \mathbf{x}_i b,$$

and furthermore

$$\theta_u^*(b) \equiv \inf_x \theta_u(x, b), \theta_l^*(b) \equiv \sup_x \theta_l(x, b).$$

Then the hypothesis $b \in B_I$ is equivalent to the hypothesis $0 \in [\theta_l^*(b), \theta_u^*(b)]$. Provided sufficient conditions for strong approximation or majorization of the suprema of studentized versions of $\hat{\theta}_l(x, b)$ and $\hat{\theta}_u(x, b)$,² one can use the results of CLR to estimate

²This is easily established with discrete covariates, but is more challenging in the continuous case. Under suitable conditions Haerdle and Song (2008) establish strong approximation of the quantile

a $1 - \alpha$ confidence interval $[\theta_l^*(b), \theta_u^*(b)]$ denoted $[\hat{\theta}_l(b), \hat{\theta}_u(b)]$ such that for any fixed b ,

$$\inf_{\theta \in [\theta_l^*(b), \theta_u^*(b)]} P \left\{ \theta \in [\hat{\theta}_l(b), \hat{\theta}_u(b)] \right\} \geq 1 - \alpha + o(1).$$

As a consequence, the hypothesis $H_0 : b \in B_I$, $H_1 : b \notin B_I$ can be conducted by rejecting if $0 \notin [\hat{\theta}_l(b), \hat{\theta}_u(b)]$ and failing to reject otherwise. The set of b such that $0 \in [\hat{\theta}_l(b), \hat{\theta}_u(b)]$ then provides a $1 - \alpha$ confidence set for β .

4 Examples

This section considers some examples to illustrate the nature of the bounds B_I . I first consider examples with $T = 2$ where the conditional distribution of the unobservables U_{i1} and U_{i2} given \mathbf{x}_i is Cauchy with location parameter zero. The Cauchy distribution is particularly convenient for computing the conditional quantiles of $\Delta U_i | \mathbf{x}_i$ necessary for construction of the identified set. Specifically, $\Delta U_i | \mathbf{x}_i$ is then also Cauchy with scale parameter the sum of those of the conditional distributions of U_{i1} and U_{i2} . In addition, the Cauchy distribution is a leading example of a distribution without moments, so that conditional mean restrictions do not hold. The identified sets illustrated make use of the assumption $Q_{U_{i1}}(\tau | \mathbf{x}_i) = Q_{U_{i2}}(\tau | \mathbf{x}_i) = 0$ for $\tau = 1/2$, in addition to Assumption WCI, but nothing more. In the figures that follow, I plot the identified set B_I derived in Theorem 2. I consider alternative specifications for the scale parameters of the conditional distributions of U_{i1} and U_{i2} given \mathbf{x}_i , denoted $\gamma_1(\mathbf{x}_i)$ and $\gamma_2(\mathbf{x}_i)$, respectively, and investigate their effect on the size and shape of the resulting identified set.³ I then consider a setting in which U_{i1} is again Cauchy, but where U_{i2} follows a shifted exponential distribution, in which case the distribution of ΔU_i is asymmetric.

Recall that two factors play a key role in determining the size of the identified set. These are the scale of the observed covariates \mathbf{X}_i and the magnitudes of the conditional quantiles $Q_{\Delta u}(\tau(1-\tau) | \mathbf{x}_i)$ and $Q_{\Delta u}(1-\tau(1-\tau) | \mathbf{x}_i)$. For illustration

regression process obtained from kernel estimation with a univariate conditioning variable. In the present context conditioning variables are necessarily multivariate, so that one would need to either generalize this result to the multivariate case, or adopt a semi-parametric approach for estimation, such as the partially linear estimator considered by Haerdle, Ritov, and Song (2009).

³Specifically, $U_{it} | \mathbf{x}_i \sim \text{Cauchy}(0, \gamma_t(\mathbf{x}_i))$, with density $g_t(u | \mathbf{x}_i) = \left(\pi \gamma_t(\mathbf{x}_i) \cdot \left[1 + \left(\frac{u}{\gamma_t(\mathbf{x}_i)} \right)^2 \right] \right)^{-1}$.

I fix the support of the covariates and consider alternative specifications for the scale parameters $\gamma_1(\mathbf{x}_i)$ and $\gamma_2(\mathbf{x}_i)$, which in turn yield different values for $Q_{\Delta u}(\tau(1-\tau)|\mathbf{x}_i)$ and $Q_{\Delta u}(1-\tau(1-\tau)|\mathbf{x}_i)$. Specifically, for each period $t \in \{1, 2\}$, it is maintained throughout that X_{it} is bivariate with finite support

$$\mathcal{X}_t = \{0, 0.5, \dots, 4\} \times \{-2, -1.75, \dots, 2\},$$

and that the support of \mathbf{X}_i is $\mathcal{X}_1 \times \mathcal{X}_2$. The set B_I is thus of a convex polytope in \mathbb{R}^2 determined as the intersection of 46,818 (i.e. 2 times the cardinality of the support of \mathbf{X}_i) linear inequality constraints on β .⁴ The true parameter value is set at $\beta = (1, 1)'$ in all examples. Note that the conditional distribution of α_i has no effect on the identified set, and is therefore left unspecified.

The first set of examples specifies that $U_{i1}|\mathbf{x}_i$ and $U_{i2}|\mathbf{x}_i$ each have constant scale parameters with respect to \mathbf{x}_i , i.e. $\gamma_1(\mathbf{x}_i) = \sigma_1$ and $\gamma_2(\mathbf{x}_i) = \sigma_2$. Figure 1 depicts identified sets obtained from all combinations with σ_1 and σ_2 each equal to either 1/2 or 1. As the figure illustrates, this is a particularly simple case. In this case all but 4 of the inequalities defining B_I are redundant, so that the identified set is the region whose boundary is given by four lines in \mathbb{R}^2 . The larger the scale parameters, the larger is the identified set. This illustrates in a particularly simple setting the more general point that the less disperse the distribution of the unobservables, the smaller is the identified set. Identical figures could also be obtained by re-scaling the support points of the covariates and keeping the scale parameters of the distribution of unobservables fixed.

Figures 2 through 4 depict more complex examples, where the scale parameters are non-trivial functions of covariates. In Figure 2 the scale parameter for the distribution of U_{i1} is given by $\gamma_1(\mathbf{x}_i) = \log(2 + |x_{i,11} + x_{i,12}|)$, and the scale parameter for the distribution of U_{i2} is given by $\gamma_2(\mathbf{x}_i) = \log(2 + |x_{i,21} + x_{i,22}|) + \sigma_1 \log(1 + |x_{i,11}|) + \sigma_2 \log(1 + |x_{i,12}|)$. The four panels of Figure 2 depict identified sets corresponding to all possible values of $(\sigma_1, \sigma_2) \in \{0, 1\}^2$. Note that when $(\sigma_1, \sigma_2) \neq (0, 0)$, the conditional distribution of U_{i2} is a function of not only x_{i2} , but also x_{i1} . The figure illustrates that a variety of shapes are possible, depending on the conditional distribution of U_{i1} and U_{i2} given \mathbf{X}_i . These sets are larger than those of Figure 1 for the simple reason that the scale parameters are both greater than or equal to 1 for all covariate values.

⁴These inequalities were used in conjunction with the Multi-Parametric Toolbox software of Kvasnica, Grieder, and Baotić (2004) to plot the resulting identified sets.

In Figure 3 the scale parameters are specified as $\gamma_1(\mathbf{x}_{i1}) = 100 \cdot \exp(-|x_{i1,1} + x_{i1,2}^2|)$, and $\gamma_2(\mathbf{x}_{i2}) = 100 \cdot \exp(-|\sigma_1 x_{i1,1}/4 + x_{i2,1} + \sigma_2 x_{i1,2}^2/10 + x_{i2,2}^2/3|)$. Again, identified sets are shown for all possible values of $(\sigma_1, \sigma_2) \in \{0, 1\}^2$, and σ_1 and σ_2 govern whether or not the scale of U_{i2} varies with $x_{i1,1}$ and $x_{i1,2}$, respectively. The identified sets are smaller than in the previous figures, driven by the fact that for some values of the covariates, the scale parameters $\gamma_1(\mathbf{x}_{i1})$ and $\gamma_2(\mathbf{x}_{i2})$ are much smaller than in the prior examples. For example, when $x_{i1,1} = 4$ and $x_{i1,2} = 2$, $\gamma_1(\mathbf{x}_{i1}) \approx 0.0335$. On the other hand, for some values of the covariates, the scale parameters are also much larger, e.g. when $x_{i1,1} = x_{i1,2} = 1$, $\gamma_1(\mathbf{x}_{i1}) \approx 13.5$. However, there are enough values of \mathbf{x}_i for which the scale of the distribution $\Delta U_i | \mathbf{x}_i$ is smaller than that of the previous examples so that when all the inequalities defining B_I are taken together, the resulting identified set is smaller. This illustrates that the scale of $\Delta U_i | \mathbf{x}_i$ need not be uniformly small over all possible \mathbf{x}_i in order to obtain tighter identified sets, but can in fact be quite large for many covariate values.

Figure 4 illustrates a similar specification to that of Figure 3, but where the functions determining $\gamma_1(\mathbf{x}_{i1})$ and $\gamma_2(\mathbf{x}_{i2})$ are no longer scaled by a factor of 100. The result is much smaller identified sets. Specifically, the scale parameters are $\gamma_1(\mathbf{x}_{i1}) = \exp(-|x_{i1,1} + x_{i1,2}^2|)$ and $\gamma_2(\mathbf{x}_{i2}) = \exp(-|\sigma_1 x_{i1,1} + x_{i2,1} + \sigma_2 x_{i1,2}^2/10 + x_{i2,2}^2|)$.

Figure 5 presents a setting in which $U_{i1} | \mathbf{x}_i$ is again distributed Cauchy, but where $U_{i2} | \mathbf{x}_i$ is exponential with parameter $\lambda(\mathbf{x}_i)$, shifted down by $\ln 2 / \lambda(\mathbf{x}_i)$ in order to have conditional median 0.⁵ As a consequence neither the distribution of $U_{i2} | \mathbf{x}_i$ nor the distribution of $\Delta U | \mathbf{x}_i$ are symmetric. To compute the identified set, I solve for the characteristic function of $\Delta U_i | \mathbf{x}_i$ and then numerically compute its 1/4 and 3/4 quantiles for each \mathbf{x}_i . In the two panels on the left side of the figure, the distribution of $U_{i2} | \mathbf{x}_i$ varies with \mathbf{x}_i , with exponential parameter $\lambda(\mathbf{x}_i) = (x_{i,21} + 1) / 5$, while in the right hand panels $\lambda(\mathbf{x}_i) = 1/2$. In the top panels the scale parameter for the distribution of $U_{i1} | \mathbf{x}_i$ is $\gamma_1(\mathbf{x}_{i1}) = \exp(-|x_{i1,1} + x_{i1,2}^2|)$, while in the bottom panels it is $\gamma_1(\mathbf{x}_{i1}) = \exp(-|x_{i1,1} + x_{i1,2}/3|)$. Unsurprisingly, the resulting identified sets are asymmetric.

⁵Specifically, $U_{i2} | \mathbf{x}_i$ has pdf $f(u; \lambda(\mathbf{x}_i)) = \lambda(\mathbf{x}_i) \exp\left\{-\lambda(\mathbf{x}_i) \cdot \left(u + \frac{\ln(2)}{\lambda(\mathbf{x}_i)}\right)\right\}$.

5 Conclusion

This paper provided novel results on the identifying power of conditional quantile restrictions in finite- T panel data models with fixed effects. The first result showed that the conditional quantile restriction by itself carries no identifying power whatsoever. I then considered the addition of a weak conditional independence restriction, and provided informative bounds on the parameter of interest when these restrictions were imposed jointly. The resulting identified set was shown to be sharp when $T = 2$, and conditions were provided that resulted in point identification of the parameter vector or some of its components. The identified set was characterized by a set of linear inequality restrictions, amenable to estimation with recently developed approaches from the set identification literature. A variety of examples were used to illustrate the nature of the identified set.

A limitation is that the assumptions of this paper did not allow for dynamics or the presence of endogenous regressors, both important considerations. Future research incorporating both of these would clearly be of interest. In principle, a similar strategy providing bounds on the quantiles of differences in within-group unobservables conditional on exogenous variables could also have identifying power in such settings.

Appendix A: Proofs

Proof of Theorem 1

Pick an arbitrary $b \in B$, and let $\mathbf{x}_i \in \mathcal{X}$. It needs to be shown that there exist random variables $(\tilde{\alpha}_i, \tilde{U}_{i1}, \dots, \tilde{U}_{iT})$ with a joint distribution $\tilde{F}(\cdot)$ conditional on \mathbf{x}_i such that (i) $Q_{\tilde{U}_{it}}(\tau|\mathbf{x}_i) = 0$, and (ii) The joint distribution of $(\tilde{Y}_{i1}, \dots, \tilde{Y}_{iT})$ given $\mathbf{X}_i = \mathbf{x}_i$ is the same as that of (Y_{i1}, \dots, Y_{iT}) given $\mathbf{X}_i = \mathbf{x}_i$, where for each t ,

$$\tilde{Y}_{it} = X_{it}b + \tilde{\alpha}_i + \tilde{U}_{it}.$$

Define

$$\tilde{U}_{it} \equiv Y_{it} - X_{it}b - \tilde{\alpha}_i,$$

where the distribution of $\tilde{\alpha}_i$ conditional on $(\mathbf{y}_i, \mathbf{x}_i, \tilde{\alpha}_i)$ is such that:

1. With probability τ , $\tilde{\alpha}_i$ is continuously distributed with respect to Lebesgue measure on $\left[\max_t \{y_{it} - x_{it}b\}, \infty\right)$. In this case $\tilde{U}_{it} \leq 0$ for all t .
2. With probability $1 - \tau$, $\tilde{\alpha}_i$ is continuously distributed with respect to Lebesgue measure on $\left(-\infty, \min_t \{y_{it} - x_{it}b\}\right)$. In this case $\tilde{U}_{it} > 0$ for all t .

Then conditional on $(\alpha_i, \mathbf{x}_i, \mathbf{y}_i)$, $\tilde{U}_{it} \leq 0$ with probability τ , and

$$\begin{aligned}
\Pr \left\{ \tilde{U}_{it} \leq 0 | x_i, \alpha_i \right\} &= \Pr \{ Y_{it} - X_{it}b - \tilde{\alpha}_i \leq 0 | \mathbf{x}_i, \alpha_i \} \\
&= \Pr \{ \tilde{\alpha}_i \geq Y_{it} - X_{it}b | \mathbf{x}_i, \alpha_i \} \\
&= \int \Pr \{ \tilde{\alpha}_i \geq Y_{it} - X_{it}b | x_i, \alpha_i, \mathbf{y}_i \} dP(\mathbf{y}_i | \mathbf{x}_i, \alpha_i) \\
&= \tau,
\end{aligned}$$

so that (i) is satisfied. (ii) is immediately satisfied by the definition of \tilde{U}_{it} .

Proof of Theorem 2

To show the “if” part of the theorem begin with equation (5) and apply the conditional quantile function for ΔU given \mathbf{x}_i to all terms, giving

$$Q_{\Delta U}(\tau(1 - \tau) | \mathbf{x}_i) \leq 0 \leq Q_{\Delta U}(1 - \tau(1 - \tau) | \mathbf{x}_i).$$

Substituting $\Delta Y - \Delta X\beta$ for ΔU then gives

$$Q_{\Delta Y_i}(\tau(1 - \tau) | \mathbf{x}_i) \leq \Delta x_i \beta \leq Q_{\Delta Y_i}(1 - \tau(1 - \tau) | \mathbf{x}_i),$$

implying that $\beta \in B_I$.

To complete the proof it must be shown that the identified set is sharp, i.e. that all $b \in B_I$ are observationally equivalent to β . Consider the events $\mathcal{D}_1 \equiv \{\Delta Xb \leq \Delta Y\}$ and $\mathcal{D}_2 \equiv \{\Delta Xb > \Delta Y\}$. To prove sharpness, conditional distributions of $\tilde{\alpha}_i | \mathbf{x}_i, \mathcal{D}_1$ and $\tilde{\alpha}_i | \mathbf{x}_i, \mathcal{D}_2$ are shown to exist such that (i) $\Pr \left\{ \tilde{U}_{i1} \leq 0 | \mathbf{x}_i \right\} = \Pr \left\{ \tilde{U}_{i2} \leq 0 | \mathbf{x}_i \right\} = \tau$, where $\tilde{U}_{it} \equiv Y_{it} - X_{it}b - \tilde{\alpha}_i$, and (ii) $\left\{ \tilde{U}_{i1} \leq 0 \right\} \perp \left\{ \tilde{U}_{i2} \leq 0 \right\} | \mathbf{x}_i$. The distribution of $\left(\tilde{\alpha}_i + X_{i1}b + \tilde{U}_{i1}, \tilde{\alpha}_i + X_{i2}b + \tilde{U}_{i2} \right)$ given $\mathbf{X}_i = \mathbf{x}_i$ then matches that of $\mathbf{Y}_i | \mathbf{x}_i$ by definition of $\tilde{\mathbf{U}}_i$.

Pick $b \in B_I$ and $\mathbf{x}_i \in \mathcal{X}$. Let $\gamma_x \equiv \Pr\{\Delta Y \leq \Delta X\beta|\mathbf{x}_i\}$. γ_x is identified and by virtue of $b \in B_I$

$$\gamma_x \in [\tau(1-\tau), 1-\tau(1-\tau)]. \quad (14)$$

Suppose \mathcal{D}_1 obtains. Then $-\infty < Y_{i1} - X_{i1}b \leq Y_{i2} - X_{i2}b < \infty$. Let the distribution of $\tilde{\alpha}_i|\mathbf{x}_i, \mathcal{D}_1$ satisfy

$$\begin{aligned} \Pr\{\tilde{\alpha}_i \geq Y_{i2} - X_{i2}b|\mathbf{x}_i, \mathcal{D}_1\} &= \delta, \\ \Pr\{Y_{i1} - x_{i1}b \leq \tilde{\alpha}_i < Y_{i2} - X_{i2}b|\mathbf{x}_i, \mathcal{D}_1\} &= \frac{\tau(1-\tau)}{\gamma_x}, \\ \Pr\{\tilde{\alpha}_i < Y_{i1} - X_{i1}b|\mathbf{x}_i, \mathcal{D}_1\} &= 1 - \delta - \frac{\tau(1-\tau)}{\gamma_x}, \end{aligned}$$

for some $\delta \in \left[0, \min\left\{\frac{\tau^2}{\gamma_x}, 1 - \frac{\tau(1-\tau)}{\gamma_x}\right\}\right]$. Now suppose instead that \mathcal{D}_2 obtains. Then $-\infty < Y_{i2} - X_{i2}b < Y_{i1} - X_{i1}b < \infty$. Let the distribution of $\tilde{\alpha}_i|\mathbf{x}_i, \mathcal{D}_2$ satisfy

$$\begin{aligned} \Pr\{\tilde{\alpha}_i \geq Y_{i1} - X_{i1}b|\mathbf{x}_i, \mathcal{D}_2\} &= \frac{\tau^2 - \delta\gamma_x}{1 - \gamma_x}, \\ \Pr\{Y_{i2} - x_{i2}b \leq \tilde{\alpha}_i < Y_{i1} - X_{i1}b|\mathbf{x}_i, \mathcal{D}_2\} &= \frac{\tau(1-\tau)}{1 - \gamma_x}, \\ \Pr\{\tilde{\alpha}_i < Y_{i2} - X_{i2}b|\mathbf{x}_i, \mathcal{D}_2\} &= 1 - \frac{\tau - \delta\gamma_x}{1 - \gamma_x}. \end{aligned}$$

Note that (14) implies that all six of the above probabilities are contained in the unit interval. Then (i) holds since

$$\begin{aligned} \Pr\{\tilde{U}_{i1} \leq 0|\mathbf{x}_i\} &= \Pr\{\tilde{\alpha}_i \geq Y_{i1} - X_{i1}b|\mathbf{x}_i\} \\ &= \gamma_x \Pr\{\tilde{\alpha}_i \geq Y_{i1} - X_{i1}b|\mathbf{x}_i, \mathcal{D}_1\} + (1 - \gamma_x) \Pr\{\tilde{\alpha}_i \geq Y_{i1} - X_{i1}b|\mathbf{x}_i, \mathcal{D}_2\} \\ &= (\delta\gamma_x + \tau(1-\tau)) + (\tau^2 - \delta\gamma_x) \\ &= \tau, \end{aligned}$$

and

$$\begin{aligned}
\Pr \left\{ \tilde{U}_{i2} \leq 0 | \mathbf{x}_i \right\} &= \Pr \left\{ \tilde{\alpha}_i \geq Y_{i2} - X_{i2}b | \mathbf{x}_i \right\} \\
&= \gamma_x \Pr \left\{ \tilde{\alpha}_i \geq Y_{i2} - X_{i2}b | \mathbf{x}_i, \mathcal{D}_1 \right\} + (1 - \gamma_x) \Pr \left\{ \tilde{\alpha}_i \geq Y_{i2} - X_{i2}b | \mathbf{x}_i, \mathcal{D}_2 \right\} \\
&= \gamma_x \delta + (1 - \gamma_x) \frac{\tau - \delta \gamma_x}{1 - \gamma_x} \\
&= \tau.
\end{aligned}$$

Condition (ii) holds since

$$\begin{aligned}
\Pr \left\{ \tilde{U}_{i1} \leq 0 \wedge \tilde{U}_{i2} \leq 0 | \mathbf{x}_i \right\} &= \left[\begin{array}{l} \gamma_x \Pr \left\{ \tilde{U}_{i1} \leq 0 \wedge \tilde{U}_{i2} \leq 0 | \mathbf{x}_i, \mathcal{D}_1 \right\} \\ + (1 - \gamma_x) \Pr \left\{ \tilde{U}_{i1} \leq 0 \wedge \tilde{U}_{i2} \leq 0 | \mathbf{x}_i, \mathcal{D}_2 \right\} \end{array} \right] \\
&= \gamma_x \Pr \left\{ \tilde{\alpha}_i \geq Y_{i2} - X_{i2}b | \mathbf{x}_i, \mathcal{D}_1 \right\} + (1 - \gamma_x) \Pr \left\{ \tilde{\alpha}_i \geq Y_{i1} - X_{i1}b | \mathbf{x}_i, \mathcal{D}_2 \right\} \\
&= \gamma_x \delta + (1 - \gamma_x) \frac{\tau^2 - \delta \gamma_x}{1 - \gamma_x} \\
&= \tau^2.
\end{aligned}$$

Proof of Corollary 1

Convexity is immediate from the linearity of the inequality restrictions. Non-emptiness follows from writing the identified set as in (9) and because $Q_{\Delta U_i}(\tau(1 - \tau) | \mathbf{x}_i) \leq 0 \leq Q_{\Delta U_i}(1 - \tau(1 - \tau) | \mathbf{x}_i)$.

Proof of Corollary 2

The proof is immediate by application of Theorem 2 to each s, t pair.

Proof of Theorem 3

Suppose that $b \neq \beta$, and consider first the case where $b_k \neq \beta_k$. Then for any positive measure bounded set of values of $\mathbf{x}_{i,-k}$,

$$\Delta x_i(b - \beta) = \Delta x_{i,k}(b_k - \beta_k) + \Delta x_{i,-k}(b_{-k} - \beta_{-k}),$$

diverges to positive or negative infinity as $\Delta x_{i,k} \rightarrow \infty$. Condition (i) and C1 then imply that one of the inequalities of (9) is violated with positive probability conditional on that set of $\mathbf{x}_{i,-k}$, from which it follows that b is identified relative to β .

Now suppose that $b \neq \beta$, but $b_k = \beta_k$. Then for all \mathbf{x}_i

$$\Delta x_i (b - \beta) = \Delta x_{i,-k} (b_{-k} - \beta_{-k}).$$

The support condition (ii) guarantees that for any $b \neq \beta$ there exists some $\delta > 0$ such that $\Pr \{\mathbf{x}_i : |\Delta x_i (b - \beta)| > \delta\} > 0$, implying that for this b in particular with $b_k = \beta_k$,

$$\Pr \{|\mathbf{x}_i : \Delta x_{i,-k} (b_{-k} - \beta_{-k})| > \delta\} > 0, \quad (15)$$

for some $\delta > 0$. Let D_b denote the set of $x_{i,-k}$ such that $|\Delta x_{i,-k} (b_{-k} - \beta_{-k})| > \delta$. Condition (i) together with the observation $Q_{\Delta U_i}(\tau(1-\tau) | \mathbf{x}_i) \leq 0 \leq Q_{\Delta U_i}(\tau(1-\tau) | \mathbf{x}_i)$ previously shown in the proof of Theorem 2 implies that there exists a constant C such that

$$-\delta < Q_{\Delta U_i}(\tau(1-\tau) | \Delta X_{i,-k} \in D_b, \Delta X_{i,k} > C) \leq 0,$$

and

$$0 \leq Q_{\Delta U_i}(1 - \tau(1 - \tau) | \Delta X_{i,-k} \in D_b, \Delta X_{i,k} > C) < \delta,$$

Combined with (15) these inequalities imply a violation of the inequalities (9). Under Assumption C1 the event $\{\Delta X_{i,k} > C\}$ has positive probability conditional on $\Delta x_{i,-k} \in D_b$. Since $\Pr \{x_{i,-k} \in D_b\} > 0$ as well, it follows that (9) is violated with positive probability and β is identified relative to b .

Proof of Theorem 4

Define the set

$$\tilde{S}(b) \equiv \{x : \Delta x(b - \beta) < C_0 \vee \Delta x(b - \beta) > C_1\}.$$

Suppose that $\mathbf{x} \in \tilde{S}(b)$, so that either $\Delta x(b - \beta) < C_0$ or $\Delta x(b - \beta) > C_1$. If $\Delta x(b - \beta) < C_0$, then it follows by C2 that $\Delta x(b - \beta) < Q_{\Delta u}(\tau(1-\tau) | \mathbf{x}_i)$. Adding $\Delta x\beta$ to both sides of the inequality gives $\Delta xb < Q_{\Delta Y}(\tau(1-\tau) | \mathbf{x}_i)$, implying that $\mathbf{x} \in S(b)$ as defined in Theorem 2. By similar reasoning, if $\Delta x(b - \beta) > C_1$ then $\Delta x(b - \beta) > Q_{\Delta u}(1 - \tau(1 - \tau) | \mathbf{x}_i)$ and $\mathbf{x} \in S(b)$. Therefore $\mathbf{x} \in \tilde{S}(b) \Rightarrow \mathbf{x} \in S(b)$.

Now re-write the set $\tilde{S}(b)$ as

$$\tilde{S}(b) = \left\{ \begin{array}{l} x : \Delta x_k (b_k - \beta_k) < C_0 - \Delta x_{-k} (b_{-k} - \beta_{-k}) \\ \vee \Delta x_k (b_k - \beta_k) > C_1 - \Delta x_{-k} (b_{-k} - \beta_{-k}) \end{array} \right\}.$$

Under condition C1 it follows that conditional on any Δx_{-k} there exists a positive measure set of Δx_k such that these inequalities hold. Therefore $\Pr \left\{ \tilde{S}(b) \right\} > 0$ and since $\mathbf{x} \in \tilde{S}(b) \Rightarrow \mathbf{x} \in S(b)$, β_k is point-identified.

References

- ABREVAYA, J. (2000): “Rank Estimation of a Generalized Fixed Effects Regression Model,” *Journal of Econometrics*, 95, 1–23.
- ABREVAYA, J., AND C. M. DAHL (2008): “The Effects of Birth Inputs on Birthweight: Evidence from Quantile Estimation on Panel Data,” *Journal of Business and Economic Statistics*, 26(4).
- ANDREWS, D. W. K., AND X. SHI (2009): “Inference for Parameters Defined by Conditional Moment Inequalities,” working paper, Cowles Foundation.
- ANDREWS, D. W. K., AND G. SOARES (2007): “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” working paper, Cowles Foundation, Yale University.
- ARELLANO, M., AND S. BONHOMME (2009): “Intersection Bounds, Estimation and Inference,” CEMMAP working paper CWP22/09.
- BERESTEANU, A., AND F. MOLINARI (2008): “Asymptotic Properties for a Class of Partially Identified Models,” *Econometrica*, 76, 763–814.
- BUGNI, F. (2007): “Bootstrap Inference in Partially Identified Models,” working paper, Northwestern University.
- CANAY, I. (2007): “EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity,” working paper, Northwestern University.
- (2008): “A Note on Quantile Regression for Panel Data Models,” working paper, Northwestern University.
- CHAMBERLAIN, G. (1982): “Multivariate Regression Models for Panel Data,” *Journal of Econometrics*, 18(1), 5–46.
- (1992): “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60(3), 567–596.
- CHAUDHURI, P. (1991): “Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation,” *Annals of Statistics*, 19(2), 760–777.

- CHERNOZHUKOV, V., I. FERNANDEZ-VAL, J. HAHN, AND W. NEWEY (2008): “Identification and Estimation of Marginal Effects in Nonlinear Panel Models,” cemmap working paper CWP05/09.
- CHERNOZHUKOV, V., AND C. HANSEN (2006): “Instrumental Quantile Regression Inference for Structural and Treatment Effect Models,” *Journal of Econometrics*, 123(2), 491–525.
- (2008): “Instrumental Variable Quantile Regression: A Robust Inference Approach,” *Journal of Econometrics*, 142(1), 379–398.
- CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): “Estimation and Confidence Regions for Parameter Sets in Econometric Models,” *Econometrica*, 75(5), 1243–1284.
- CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2009): “Intersection Bounds, Estimation and Inference,” CEMMAP working paper CWP19/09.
- GALICHON, A., AND M. HENRY (2009): “A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters,” *Journal of Econometrics*, forthcoming.
- GALVAO, A. (2008): “Quantile Regression for Dynamic Panel Data with Fixed Effects,” working paper, University of Illinois, Urbana-Champaign.
- GERACI, M., AND M. BOTTAI (2007): “Quantile Regression for Longitudinal Data Using the Asymmetric Laplace Distribution,” *Biostatistics*, 8(1), 140–154.
- GRAHAM, B. S., J. HAHN, AND J. L. POWELL (2008): “The Incidental Parameter Problem in a Non-Differentiable Panel Data Model,” working paper, University of California, Berkeley.
- (2009): “A Quantile Correlated Random Coefficient Panel Data Model,” working paper, University of California, Berkeley.
- GRAHAM, B. S., AND J. L. POWELL (2008): “Identification and Estimation of ‘Irregular’ Correlated Random Coefficient Models,” working paper, University of California, Berkeley.

- HAERDLE, W., Y. RITOV, AND S. SONG (2009): “Partial Linear Quantile Regression and Bootstrap Confidence Bands,” working paper, Humboldt University.
- HAERDLE, W., AND S. SONG (2008): “The Stochastic Fluctuation of the Quantile Regression Curve,” SFB 649 Discussion Paper 2008-027.
- HAN, A. K. (1987): “Non-parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator,” *Journal of Econometrics*, 35, 303–316.
- HARDING, M. C., AND C. LAMARCHE (2008): “A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables,” working paper, Stanford University.
- HONORE, B. E. (1992): “Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects,” *Econometrica*, 60(3), 533–565.
- HONORE, B. E., AND E. TAMER (2006): “Bounds on Parameters in Panel Dynamic Discrete Choice Models,” *Econometrica*, 74(3), 611–629.
- KOENKER, R. (2004): “Quantile Regression for Longitudinal Data,” *Journal of Multivariate Analysis*, 91, 74–89.
- (2005): *Quantile Regression*. Cambridge University Press, Cambridge.
- KOENKER, R., AND G. BASSETT (1978): “Regression Quantiles,” *Econometrica*, 46(1), 33–50.
- KOENKER, R., AND K. F. HALLOCK (2000): “Quantile Regression: An Introduction,” working paper, University of Illinois, Urbana Champaign.
- KOENKER, R., P. NG, AND S. PORTNOY (1994): “Quantile Smoothing Splines,” *Biometrika*, 81(4), 673–680.
- KVASNICA, M., P. GRIEDER, AND M. BAOTIĆ (2004): “Multi-Parametric Toolbox (MPT),” <http://control.ee.ethz.ch/mpt/>.
- LAMARCHE, C. (2006): “Robust Penalized Quantile Regression for Panel Data,” working paper, University of Oklahoma.

- (2008): “Penalized Quantile Regression Estimation for a Model with Endogenous Individual Effects,” working paper, University of Oklahoma.
- MANSKI, C. F. (1985): “Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator,” *Journal of Econometrics*, 27, 313–333.
- (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(2), 519–546.
- MENZEL, K. (2009): “Estimation and Inference with Many Weak Moment Inequalities,” working paper, MIT.
- ROMANO, J. P., AND A. M. SHAIKH (2008): “Inference for Identifiable Parameters in Partially Identified Econometric Models,” *Journal of Planning and Statistical Inference*, 138, 2786–2807.
- ROSEN, A. M. (2008): “Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities,” *Journal of Econometrics*, 146, 107–117.
- WOOLDRIDGE, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

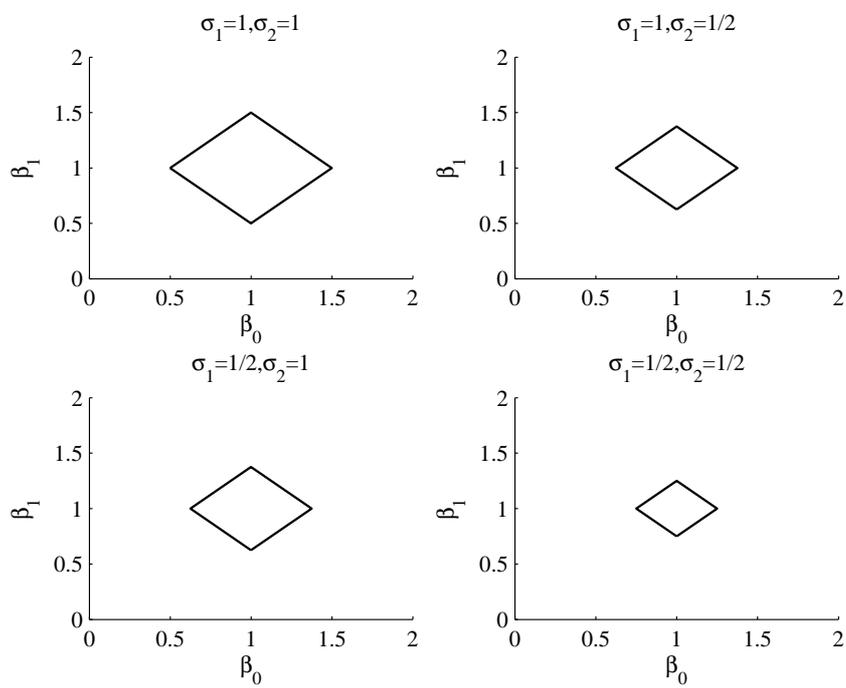


Figure 1: The identified set when the conditional distributions of U_{i1} and U_{i2} are Cauchy with scale parameters σ_1 and σ_2 .

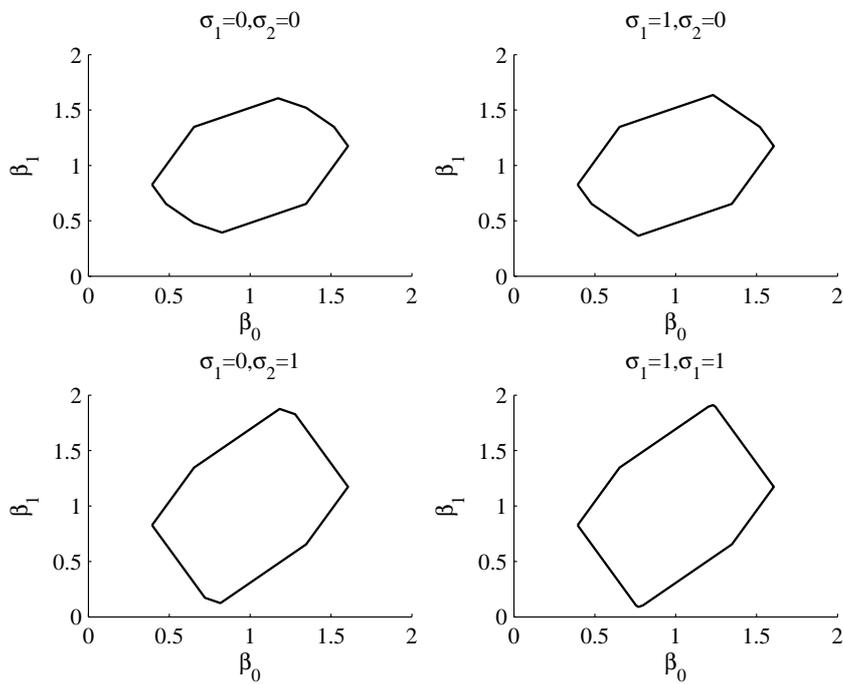


Figure 2: The identified set when the conditional distributions of U_{i1} and U_{i2} are Cauchy with scale parameters $\gamma_1 = \log(2 + |x_{i1,1} + x_{i1,2}|)$ and $\gamma_2 = \log(2 + |x_{i2,1} + x_{i2,2}|) + \sigma_1 \log(1 + |x_{i1,1}|) + \sigma_2 \log(1 + |x_{i1,2}|)$.

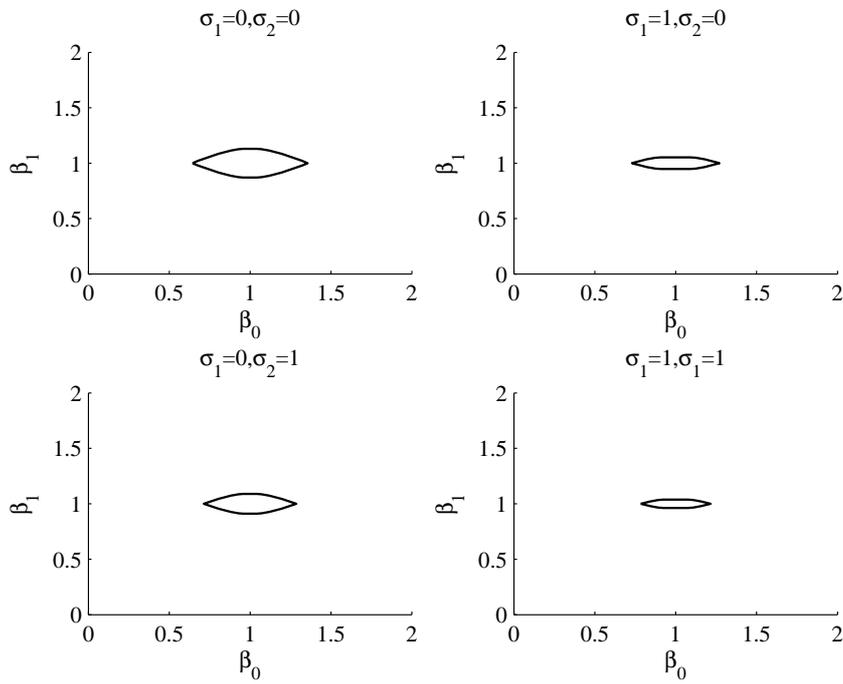


Figure 3: The identified set when the conditional distributions of U_{i1} and U_{i2} are Cauchy with scale parameters $\gamma_1 = 100 \exp\{-|x_{i1,1} + x_{i1,2}|\}$ and $\gamma_2 = 100 \exp\left\{-\left|\sigma_1 \frac{x_{i1,1}}{4} + x_{i2,1} + \sigma_2 \frac{x_{i1,2}^2}{10} + \frac{x_{i2,2}^2}{3}\right|\right\}$.

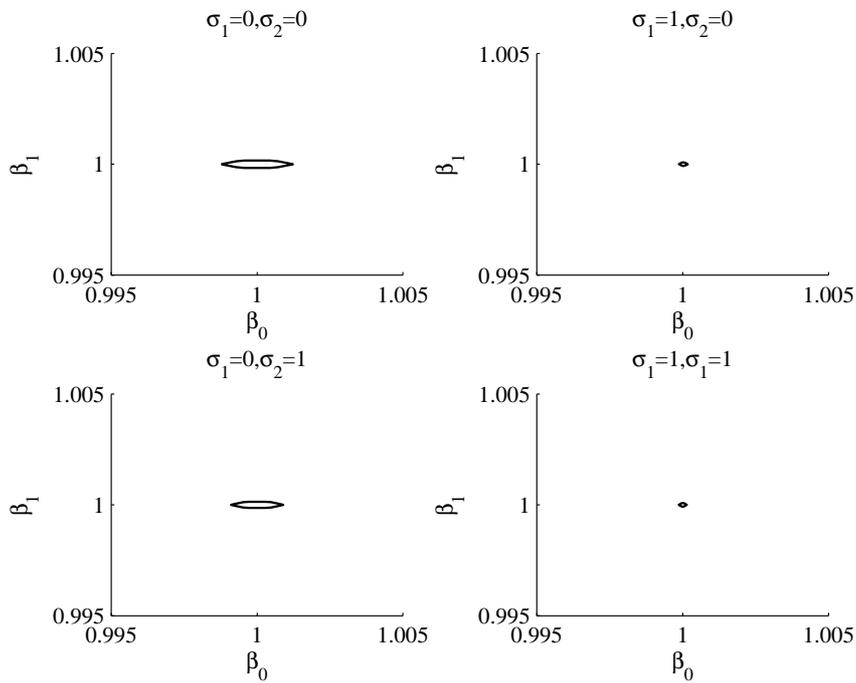


Figure 4: The identified set when the conditional distributions of U_{i1} and U_{i2} are Cauchy with scale parameters $\gamma_1(\mathbf{x}_{i1}) = \exp(-|x_{i1,1} + x_{i1,2}^2|)$ and $\gamma_2(\mathbf{x}_{i2}) = \exp(-|\sigma_1 x_{i1,1} + x_{i2,1} + \sigma_2 x_{i1,2}^2/10 + x_{i2,2}^2|)$.

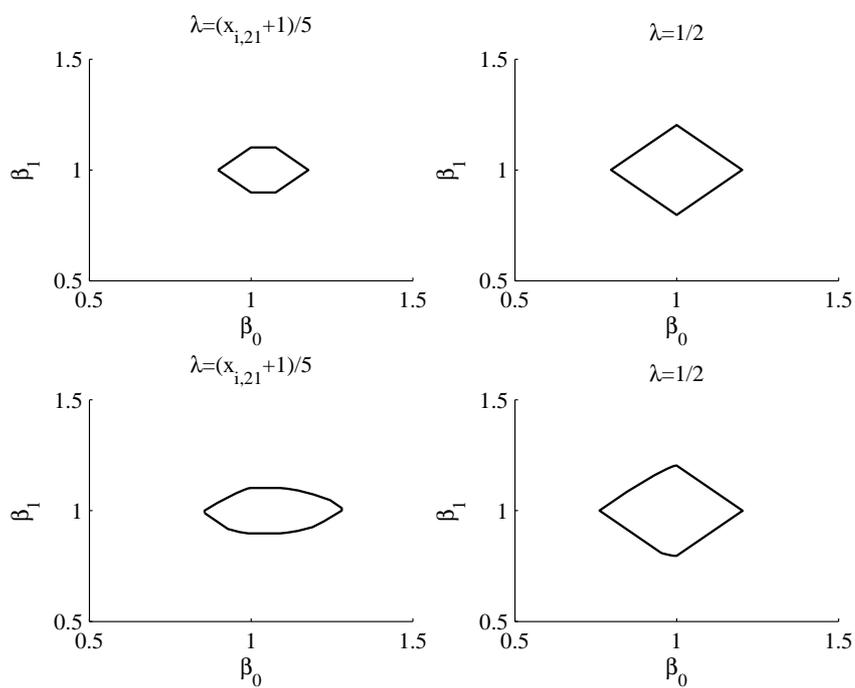


Figure 5: The identified set when the conditional distributions of U_{i1} and U_{i2} are Cauchy and shifted exponential, respectively.