Identified Regions and Inference in Panel Data Roy Models

Shakeeb Khan\textsuperscript{a} and Elie Tamer\textsuperscript{b}

\textsuperscript{a}Department of Economics, Duke University, Durham, NC 27708.
\textsuperscript{b}Department of Economics, Northwestern University, 2001 Sheridan Rd, Evanston, IL 60208.

August 2009
Preliminary and Incomplete Version:
Please do not cite or circulate.

Abstract

This paper explores the identifiability of regression coefficients in a panel data Roy model. Our approach is based on deriving the form of \textit{conditional moment inequalities} which can be used to infer the identified region of the parameter space. Our set of moment inequalities is complete in the sense the bounds we attain the parameters are sharp. A method based on a transformation to what we call \textit{unconditional integrated moment inequalities} is proposed, to which sample analogs can be used to consistently estimate the identified region. Sufficient conditions for point identification and asymptotic properties of the proposed estimation procedure are derived. An extension to identification in cross sectional Roy models is considered, and an inference method based on conditional moment inequalities after \textit{pairwise differencing} is proposed. Finite sample properties of the proposed methods are explore through a small scale simulation study.

\textsuperscript{*}Corresponding Author: Department of Economics, Duke University, 213 Social Sciences Building, Durham, NC 27708. Phone: (919) 660-1873.
1 Introduction

The Roy model Roy (1951) and its extensions has received much attention in both the applied and theoretical econometrics literature. For a recent survey on this literature, see Heckman and Vytlacil (2008). The large and growing literature is undoubtedly due to the fact that Roy’s framework, developed to explain occupational choice and its effect on the distribution of earnings, is but one example for an entire class of models of self selection in markets. Most, but not all markets considered in the literature, relate in some way to labor markets- examples include choice between market and nonmarket work and the consequence on observed wages Gronau (1974), Heckman (1978), a worker’s choice between union and nonunion sectors Lee (1978), industrial sectors Heckman and Sedlaceck (1990), and their effects on earnings all fall within the general framework of the Roy model.

Paralleling this burgeoning literature in econometrics is the growing literature in panel data models. The increased availability of longitudinal panel data sets has presented new opportunities for econometricians to control for individual unobserved heterogeneity across agents. Panel data is related to grouped data sets, where in this context group specific heterogeneity can be controlled for.

In linear panel data models, unobserved additive individual specific heterogeneity, if assumed constant over time or group member (i.e. “fixed effects”), can be controlled for when estimating the slope parameters by first differencing the observations. Unfortunately, the first differencing approach will typically be inapplicable for estimating nonlinear models, such as the Roy model.

The lack of adequate identification results for the Roy model with the unobserved heterogeneity in panel data models is precisely what motivates the work in this paper. Specifically, we aim to explore the identified regions of parameters of interest in a Roy model with the structure of unobserved heterogeneity usually found in nonlinear panel data models. In one sense what we do in this paper can be viewed as a a generalization of the seminal Heckman and Honoré (1990) paper, who wished to study the empirical content of the Roy model. In that paper the authors wished to determine under which conditions the parameters of interest could be identified when relaxing the normality assumption in the original Roy model. The results here are a generalization in two ways. For one, our model will allow for individual specific effects that are meant to allow for further unobserved heterogeneity. Second, we aim to pay attention to sharp identified regions, as opposed to the dichotomous identification/identification results in Heckman and Honoré (1990).
In that sense our results aim to add to the growing literature on partial identification. (see Manski (2003) for a recent survey on this literature.) Our approach for constructing and estimating the identified regions will be based on conditional moment inequalities. See Khan and Tamer (2009) and references therein for recent work in this growing area in econometrics. The rest of the paper is organized as follows. The next section introduces the basic model whose parameters we wish to conduct inference, and states the assumptions the ensuing identification results will be based on. We will then propose the conditional moment inequalities, and what we will refer to as integrated moment inequalities for the parameters of interest that will serve as the basis for constructing estimation and inference. This will suggest will propose (set) estimation procedures which are based on sample analogs of the derived moment inequalities, an establish their asymptotic properties for the purpose of conducting inference. A subsection extends the ideas by proposing a cross sectional analog based on moment inequalities obtained after pairwise differencing, to which identification results can be compared to Heckman and Honoré (1990). Finally, section 3 concludes by proposing further extensions for future research.

2 Two Sector Panel Data Model

To illustrate the identification approach taken in this paper we first introduce the two sector 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:

\[ y_{it}^* = \alpha_i + x_{it}' \beta_0 + \epsilon_{it} \]  

where \( i = 1, 2, \ldots N, \ t = 1, 2 \). \( \alpha_i \) is an unobserved individual specific “fixed” effect, and we assume the unobserved disturbance terms \( \epsilon_{i1}, \epsilon_{i2} \) are strictly stationary given \( x_i \equiv x_{i1}, x_{i2} \) and \( \alpha_i \). As discussed in Honoré and Kyriazidou (2000), the strict stationarity assumption generalizes the conditional exchangeability assumption in Honoré (1992) which itself is more general than an i.i.d assumption. Nonetheless, to illustrate our basic identification strategy, we will assume \( \epsilon_{i1}, \epsilon_{i2} \) are i.i.d., conditional on \( x_i, \alpha_i \), stating here that it is stronger than necessary, and can easily be relaxed to a conditional stationarity 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 \( N \).

To exposit the two-sector Roy model, here we do not always observe \( y_{it}^* \). Instead, the
econometrician observes \( v_{it} \equiv \max(y^*_{it}, s_{it}) \), where \( s_{it} \) 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 \( s_{it} \) 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 correlated with \( x_i, \alpha_i, \epsilon_{it} \).

Of interest will be identification and inference on the regression coefficients \( \beta_0 \), and our approach will be based on conditional moment inequalities.

To construct the conditional moment inequalities our inference will be based on, we first construct the following observable random variables from our current observed dependent variables:

\[
y^1_{it} = v_{it} \tag{2.2}
\]

and

\[
y^0_{it} = d_{it}v_{it} + (1 - d_{it}) \cdot (-\infty) \tag{2.3}
\]

From these definitions we immediately have the following inequalities:

\[
y^0_{it} \leq y^*_{it} \leq y^1_{it} \tag{2.4}
\]

Note since this holds for \( t = 1, 2 \) this will imply the following two inequalities:

\[
y^0_{i1} - y^1_{i2} \leq y^*_{i1} - y^*_{i2} \tag{2.5}
\]

\[
y^*_{i1} - y^*_{i2} \leq y^1_{i1} - y^0_{i2} \tag{2.6}
\]

Note the LHS of the above equation is simply

\[
\Delta x'_i \beta_0 + \Delta \epsilon_i
\]

where \( \Delta \) denotes the time difference operator. Note also, that given our error terms assumption, \( \Delta \epsilon_i \) is conditionally (on \( \alpha_i, x_i \)) distributed symmetrically around 0.

We further note one can write the inequality in (2.6) as:

\[
\Delta \epsilon_i \leq y^1_{i1} - y^0_{i2} - \Delta x'_i \beta_0 \tag{2.7}
\]
which implies the following conditional moment inequality:

\[
\frac{1}{2} = P(\Delta \varepsilon_i \geq 0 | x_i) \leq P(y_{i1}^1 - y_{i2}^0 - \Delta x_i' \beta_0 \geq 0 | x_i) \quad (2.8)
\]

Using the same arguments, we also have the following conditional moment inequality:

\[
\frac{1}{2} = P(\Delta \varepsilon_i \leq 0 | x_i) \geq P(y_{i1}^0 - y_{i2}^1 - \Delta x_i' \beta_0 \geq 0 | x_i) \quad (2.9)
\]

This now becomes loosely analogous to the setup in Khan and Tamer (2009)- specifically, a set of two conditional moment inequalities. It is important to note that these inequalities were shown using the “true” (i.e. d.g.p. generating) parameter vector \( \beta_0 \), and importantly, may not hold for \( \beta \neq \beta_0 \). Hence these inequalities will the basis for our definition of the identified region, to be made precise below. Furthermore, we point out that the above righthand side terms involve observable variables \( y_{it}^0, y_{it}^1 \) so we can replace the moments with sample analogs when it comes to estimation and inference.

However, for the model at hand, the conditional symmetry of \( \Delta \varepsilon_i \) will introduce an additional set of conditional moment inequalities unavailable in Khan and Tamer (2009), which can further shrink the identified region. Note for any value \( \tau \in R \), we have:

\[
1 = P(\Delta \varepsilon_i \geq \tau | x_i) + P(\Delta \varepsilon_i \geq -\tau | x_i, \alpha_i) \quad (2.10)
\]

Which implies the following conditional moment inequality:

\[
1 \leq P(y_{i1}^1 - y_{i2}^0 - \Delta x_i' \beta_0 \geq \tau | x_i, \alpha_i) + P(y_{i1}^0 - y_{i2}^1 - \Delta x_i' \beta_0 \geq -\tau | x_i, \alpha_i) \quad (2.11)
\]

Which will be denoted as:

\[
1 \leq E[\psi_1(y_{i1}^1, y_{i2}^0, \beta_0, \tau) | x_i] \quad (2.12)
\]

and similarly, we have, defining \( \psi_2 \) analogous to the way we define \( \psi_1 \), :

\[
1 \geq E[\psi_2(y_{i1}^1, y_{i2}^0, \tau, \beta_0 | x_i)] \quad (2.13)
\]

To simplify notation further, we will express the above inequalities as:

\[
0 \leq g_1(\beta_0, \tau, x_i) \quad (2.14)
\]

\[
0 \geq g_2(\beta_0, \tau, x_i) \quad (2.15)
\]

These conditional moment inequalities will be our first basis for identifying (point or set) \( \beta_0 \). From the above inequalities we can define the identified region as follows:
Definition 1 Assume $\beta_0 \in \mathcal{B}$, a compact parameter space. We define the identified region as the subset of $\mathcal{B}$, denoted by $\mathcal{B}_I$, as the set of values of $\beta$ such that:

\begin{align}
0 &\leq g_1(\beta, \tau, x_i) \\
0 &\geq g_2(\beta, \tau, x_i)
\end{align} (2.16)

for all values of $x_i$ in its support and for all values of $\tau \in \mathbb{R}$.

Thus the key for identification is that for a value $\tilde{\beta}$ not in the identified region, there will exist a value of $x_i$ and a value of $\tau$ such that $g_1(\tilde{\beta}, \tau, x_i) < 0$ or $g_2(\tilde{\beta}, \tau, x_i) > 0$.

From an estimation point of view, i.e. consistently estimating the identified region, our next step is to transform conditional moment inequalities into unconditional moment inequalities. Khan and Tamer (2009), transformed conditional moment inequalities to unconditional moment inequalities using third order $U$ statistics, comparing regressor values of cross sectional units. However, that is not possible in our panel data setting, where, due to the individual specific effects $\alpha_i$, the distributions of the regressors can vary across individuals under our assumptions. Nonetheless, we can transform the conditional moment inequalities as follows. Let $d(x_i)$ denote a function of $x_i$ whose dimension is greater or equal to that of $x_i$ and for which each component is positive, e.g. $d(x_i) = |x_{i2} - x_{i1}|$. See, for example Andrews and Shi (2009) for restrictions on $d()$ so no identifying power is lost when transforming to unconditional inequalities.

From this we can define the following functions:

$$
\tilde{H}_1(\beta, \tau, d(x_i)) = E[(\psi_1(y_{i1}^1, y_{i2}^0, \beta, \tau) - 1)d(x_i)]
$$

and

$$
\tilde{H}_0(\beta, \tau, d(x_i)) = E[(\psi_2(y_{i1}^1, y_{i2}^0, \beta, \tau) - 1)d(x_i)]
$$

With these definitions we can define our population objective function as

$$
Q(\beta) = E[\omega(\tau)\tilde{H}_0(\beta, \tau, d(x_i))I[\tilde{H}_0(\beta, \tau, d(x_i)) \geq 0]]
$$

$$
- E[\omega(\tau)\tilde{H}_1(\beta, \tau, d(x_i))I[\tilde{H}_1(\beta, \tau, x_i) \leq 0]]
$$
where the above expectations is taken across \(x_i, \tau\), and \(\omega(\tau)\) is a weighting function, that is nonnegative, and \(\int w(\tau)d\tau = 1\). Thus we can see we are effectively integrating conditional moment conditions, analogous to Khan and Tamer (2009) and ?, but we here we also integrating with respect to quantiles \(\tau\).

Note the above expectations are 0 at \(\beta = \beta_0\), and their difference is greater or equal to 0 otherwise. Thus from this setup we can alternatively define the identified region as the values of the \(\beta\) where \(Q(\beta) = 0\).

We can now define our (point or set) estimator of the identified region as the minimizer(s) of a sample analog objective function. To do so we first define the functions:

\[
\tilde{H}_{1n}(\beta, x_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{m=1}^{T} (\psi_1(y_{i1}^1, y_{i2}^0, \beta, \tau_m) - 1)d(x_i)
\]

and

\[
\tilde{H}_{0n}(\beta, x_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T} \sum_{m=1}^{T} (\psi_2(y_{i1}^1, y_{i2}^0, \beta, \tau_m) - 1)d(x_i)
\]

where here \(\tau_m\) denotes \(T\) “moment points” providing us with additional moment conditions.

From which we can define our estimator as:

\[
\hat{\beta} = \arg \min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} \tilde{H}_{0n}(\beta, x_i)I[\tilde{H}_{0n}(\beta, x_i) \geq 0] - \frac{1}{n} \sum_{i=1}^{n} \tilde{H}_{1n}(\beta, x_i)I[\tilde{H}_{1n}(\beta, x_i) \leq 0]
\]

We next turn attention to the asymptotic properties of the estimator. Our first result will be to establish consistency of the estimator, beginning with a standard point consistency theorem. Our theorem is based on the following assumptions, which are similar to those in Honoré (1992):

\begin{itemize}
  \item \textbf{P1} The parameter space \(\mathcal{B}\) is compact.
  \item \textbf{P2} The conditional distribution of \(\epsilon_{i1}, \epsilon_{i2}\), given \(\alpha_i, x_{i1}, x_{i2}\) is continuous with finite density.
  \item \textbf{P3} The distribution of \(\epsilon_{i1} - \epsilon_{i2}\) conditional on \(\alpha_i, x_{i1}, x_{i2}\) is symmetric around 0.
\end{itemize}
There is no proper linear subspace of $\mathbb{R}^K$ containing the random variable $E[d_1d_2|x_{i1}, x_{i2}](x_{i2} - x_{i1})$ with probability one.

The limiting objective function

$$E[\tilde{H}_0(\beta, x_i)I[\tilde{H}_{0n}(\beta, x_i) \geq 0]$$

$$-E[\tilde{H}_1(\beta, x_i)I[\tilde{H}_{1n}(\beta, x_i) \leq 0]]$$

has a unique minimum at $\beta = \beta_0$.

Remark 1 Assumption P5 is a high level assumption for which interpretable sufficient conditions can be provided. For example one set of conditions would be the usual full rank condition on the covariates if one included "extreme" quantiles in the set of moment inequalities. While this would call into question the "regularity" of the identification condition in A2, in the following section we illustrate how point identification can be attained without resorting to extreme quantiles.

These conditions suffice for the (point) consistency of our estimator:

Theorem 2.1 Under Assumptions P1-P5,

$$\hat{\beta} \overset{p}{\to} \beta_0 \quad (2.18)$$

2.0.1 Point and Set Inference (to be completed)

Inference can be conducted using similar arguments and conditions to those previously considered in the literature. For example, when we do indeed have point identification limiting distribution theory of the estimator will follow from the same conditions and arguments used in Khan and Tamer (2009); otherwise objective function based set inference procedures such as Chernozhukov, Hong, and Tamer (2007),Kim (2008),Andrews and Shi (2009) can be modified for the problem at hand. The details of such extensions are to be completed.
2.1 Extension: Inference in a Cross Sectional Model using Pair-wise Comparisons and additional Moment Points

Here we consider the following cross-sectional model:

\[ y_i^* = x_i' \beta_0 + \epsilon_i \]

where \( i = 1, 2, \ldots, n \). Here \( y_i^* \) is unobserved as is \( \epsilon_i \), which we assume here to be distributed independently of the observed vector \( x_i \).

To fit into our framework, one observes \( v_i \equiv \max(y_i^*, c_i) \), where \( c_i \) denotes the wage offered in a different sector, and the indicator \( d_i \) which denotes which sector the wage is drawn from. Of interest again is set identification and inference on the vector of regression coefficients \( \beta_0 \). As in the panel data setting, we impose no structure in this sector, allowing \( c_i \) to correlated with \( x_i, \epsilon_i \). This again allows to consider randomly censored regression models as special cases. Therefore, related work would include the recent literature on randomly censored and competing risks models- see, e.g. Khan and Tamer (2009) for a list of references in the econometrics and biostatistics literature.

Under the weaker conditional median assumption on \( \epsilon_i \), Khan and Tamer (2009) propose a procedure for conducting such inference. The question aimed to be addressed here is how much tighter the identified set can become when strengthening the assumption to independence between the errors and the covariates. Related to this point, of interest will be under what conditions can point identification be achieved, and how does that relate to the support conditions imposed in Khan and Tamer (2009).

We first answer these question by an approach analogous to procedure discussed in the previous section, now involving pairs of individuals in the cross section as opposed to the pair of observations for the same individual in the panel. To see how, we define the following variables:

\[ y_i^1 = v_i \quad (2.19) \]

and

\[ y_i^0 = d_i v_i + (1 - d_i) \cdot (-\infty) \quad (2.20) \]
Note the following inequalities:
\[ y_i^0 \leq y_i^* \leq y_i^1 \quad (2.21) \]

Note since this holds for the pair of observations \( i, j \) this will imply the following two inequalities:
\[ y_i^0 - y_j^1 \leq y_i^* - y_j^* \quad (2.22) \]
\[ y_i^* - y_j^* \leq y_i^1 - y_j^0 \quad (2.23) \]

Note the LHS of the above equation is simply
\[ \Delta x_i^\prime \beta_0 + \Delta \epsilon_i \]
where now \( \Delta \) denotes the “pairwise” difference operator- see, e.g. Honoré and Powell (1994).
So by analogous arguments to those used in the previous section, we have the following conditional moment inequality for any \( \tau \in R \):
\[ 1 \leq P(y_i^1 - y_j^0 - \Delta x_i^\prime \beta_0 \geq \tau | x_i, x_j) + P(y_i^1 - y_j^0 - \Delta x_i^\prime \beta_0 \geq -\tau | x_i, x_j) \quad (2.24) \]
\[ 1 \geq P(y_i^0 - y_j^1 - \Delta x_i^\prime \beta_0 \geq \tau | x_i, x_j) + P(y_i^0 - y_j^1 - \Delta x_i^\prime \beta_0 \geq -\tau | x_i, x_j) \quad (2.25) \]

where we now have used the assumption that \( \epsilon_i, \epsilon_j \) are i.i.d., and thus their difference is symmetrically distributed around 0.

Which will be denoted as:
\[ 1 \leq E[\psi_1(y_i^1, y_j^0, \beta_0, \tau) | x_i, x_j] \quad (2.26) \]
and similarly, we have:
\[ 1 \geq E[\psi_2(y_i^1, y_j^0, \tau, \beta_0 | x_i, x_j)] \quad (2.27) \]

These will be satisfied for all values of \( \tau \) and all regressor values when \( \beta = \beta_0 \). Consequently we can define the identified region as the set of values \( \beta \) where the above inequalities hold for all values of \( \tau, x_i, x_j \).
Finally, we can estimate this set analogously to way proposed before—replace expectations with averages and convert conditional moment inequalities to unconditional moment inequalities. However, we are now able to use additional cross sectional regressor values, say \(x_l, x_m\) to construct an objective function without the need of any nonparametric methods and associated smoothing parameters.

Note, however that the conversion to unconditional moment inequalities can now be based on the \(U\)-statistic approach in Khan and Tamer (2009) since the moment condition is no longer based on conditioning on an individual specific effect.

We can outline this procedure here; first, define the functions:

\[
\hat{H}_1(\beta, \tau, x_l, x_m) = \frac{1}{n(n-1)} \sum_{i \neq j} (1 - I[y_i^1 - y_j^0 - \Delta x_i' \beta \geq \tau])I[x_l \leq x_i \leq x_m]I[x_l \leq x_j \leq x_m]
\]

\[
\hat{H}_0(\beta, \tau, x_l, x_m) = \frac{1}{n(n-1)} \sum_{i \neq j} (1 - I[y_i^0 - y_j^0 - \Delta x_i' \beta \geq \tau])I[x_l \leq x_i \leq x_m]I[x_l \leq x_j \leq x_m]
\]

From which we can define our objective function as:

\[
Q_n(\beta) = \frac{1}{n(n-1)(n-2)} \sum_{i \neq l \neq m} -\omega(\tau_i)\hat{H}_0(\beta, \tau_i, x_l, x_m)I[\hat{H}_0(\beta, \tau_i, x_l, x_m) \leq 0] \\
+ \omega(\tau_i)\hat{H}_1(\beta, \tau_i, x_l, x_m)I[\hat{H}_1(\beta, \tau_i, x_l, x_m) \geq 0]
\]

where \(\tau_i\) denotes a set of grid points from which the moment inequalities are constructed, and \(\omega(\cdot)\) is the weight function discussed in the previous section.

Therefore, \(\hat{\beta}\) can be the set of values that minimize \(Q_n(\beta)\). Having defined the procedure, we next turn attention to exploring its asymptotic properties. As the procedure is of a similar structure to those considered in Khan and Tamer - Khan and Tamer (2005) and Khan and Tamer (2009), we simply state here sufficient conditions for consistency:

The vector \((v_i, d_i, x_i, \epsilon_i)\) is i.i.d.
A2 The function
\[
Q_0(\beta) = E[-\omega(\tau) \hat{H}_1(\beta, \tau, x_l, x_m) I[\hat{H}_1(\beta, \tau, x_l, x_m) \leq 0] + \omega(\tau) \hat{H}_0(\beta, \tau, x_l, x_m) I[\hat{H}_0(\beta, \tau, x_l, x_m) \geq 0]]
\] (2.31)

has a unique maximum at $\beta = \beta_0$.

A3 The matrix $E[I[x_i, x_j \in \mathcal{C}] \Delta_{ij} x \Delta_{ij} x']$ is of full rank, where
\[
\mathcal{C} = \{x_i, x_j \in \mathcal{X} \times \mathcal{X} : P(d_i = 1|x_i) > 0, P(d_j = 1|x_j) > 0\}
\]

A4 The regressors $x_i$ are assumed to have compact support, denoted by $\mathcal{X}$, and are continuously distributed with a density function $f_X(\cdot)$ that is bounded away from 0 on $\mathcal{X}$.

A5 The error terms $\epsilon_i$ are distributed independently of $x_i$ and are continuously distributed with density function $f_\epsilon(\epsilon)$ which is bounded above.

A6 $\beta_0 \in \mathcal{B}$, a compact set.

Remark 2 Assumption A2 is a high level assumption analogous to PAgain, interpretable sufficient conditions can be provided, such as full rank condition on pairwise differenced covariates and including "extreme" quantiles in the set of moment inequalities. As we will see below, the introduction of "artificial" censoring points will prove useful in avoiding the use of extreme quantiles.

From the above conditions, pretty standard in the literature, we have the following theorem:

Theorem 2.2 Under Assumptions A1-A6 above, we have:
\[
\hat{\beta} \overset{P}{\to} \beta_0
\] (2.32)

2.2 Sharper Regions with Additional Censoring Points

In this subsection, we illustrate an alternative set estimation procedure, that has the the advantage of potentially resulting in sharper identification regions in certain censoring structures, but at the computation expense of requiring an additional "moment point"- such additional conditional moment inequalities can be attained by artificially censoring the data.
To illustrate the procedure, we will keep notation as before, working, with $y_i^0, y_j^1, x_i, x_j$.

Recall our procedure was based on comparing $i, j$ “residual pairs”:

$$e_i^0 = y_i^0 - x'_i \beta_0$$  \hspace{1cm} (2.33)  
$$e_j^1 = y_j^1 - x'_j \beta_0$$  \hspace{1cm} (2.34)

This procedure described here will be based on the censoring pair $\kappa_1, \kappa_2$, from which we can construct ”censored residual pairs”:

$$e_{ij}^0(\kappa_1, \kappa_2) = \max(y_i^0 - x'_i \beta_0, \kappa_1, \kappa_2)$$  \hspace{1cm} (2.35)  
$$e_{ij}^1(\kappa_1, \kappa_2) = \max(y_j^1 - x'_j \beta_0, \kappa_1, \kappa_2)$$  \hspace{1cm} (2.36)

We note that any such censoring preserves the inequalities discussed above, and hence we can work with conditional moment inequalities based on the censored residual pair.

To illustrate the objective function, let

$$e_{ij}^0(\beta, \kappa_1, \kappa_2) = \max(y_i^0 - x'_i \beta, \kappa_1, \kappa_2)$$  \hspace{1cm} (2.37)  
$$e_{ij}^1(\beta, \kappa_1, \kappa_2) = \max(y_j^1 - x'_j \beta, \kappa_1, \kappa_2)$$  \hspace{1cm} (2.38)

Now we can define the functions:

$$\hat{H}_1^c(\beta, \kappa, x_l, x_m) = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{s \neq t} \left( \frac{1}{2} - I[e_{ij}^1(\beta, \kappa_s, \kappa_t) \geq e_{ji}^0(\beta, \kappa_s, \kappa_t)] \right) I[x_l \leq x_i \leq x_m] I[x_l \leq x_j \leq x_m]$$  \hspace{1cm} (2.39)

and

$$\hat{H}_0^c(\beta, \kappa, x_l, x_m) = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{s \neq t} \left( \frac{1}{2} - I[e_{ij}^0(\beta, \kappa_s, \kappa_t) \geq e_{ji}^1(\beta, \kappa_s, \kappa_t)] \right) I[x_l \leq x_i \leq x_m] I[x_l \leq x_j \leq x_m]$$  \hspace{1cm} (2.40)

where $\kappa$ denotes the vector of artificial censoring points, whose dimension is at least the number of cross sectional observations.

From which we can define our new objective function as:
\[ Q_n(\beta) = \frac{1}{n(n-1)(n-2)} \sum_{i \neq l \neq m} \hat{H}_c^i(\beta, \kappa, x_l, x_m)I[\hat{H}_c^i(\beta, \kappa, x_l, x_m) \geq 0] - \hat{H}_0^i(\beta, \kappa, x_l, x_m)I[\hat{H}_0^i(\beta, \kappa, x_l, x_m) \leq 0] \] (2.41)

Note the here we only worked with one value of \( \tau = 0 \). This was only done to simplify notation, and as in previous sections a grid of \( \tau \)'s could be used. Therefore, this objective function will be based on integrating our conditional moment inequalities with respect to regressors as in Khan and Tamer (2009), Andrews and Shi (2009), and also quantiles \( \tau \) and censoring values \( \kappa \); these additional two variables aim to reduce the size of the identified set and possibly improve the power of objective function based set inference procedures—see, e.g. Chernozhukov, Hong, and Tamer (2007), Kim (2008) and Andrews and Shi (2009).

The identified region can again be defined as the values of \( \beta \) which set the above objective function to 0. As we’ll see, particular values of \( \kappa \) can result in point identification.

### 2.2.1 Attaining Point Identification

We can illustrate here how point identification can be achieved for particular values of \( \kappa \) in particular cases; this will illustrate the advantage of our moment inequality approach over kaplan meier weighted methods, which generally cannot accommodate dependent risks, and generally require smoothing parameters for conditionally independent risks.

Point identification will be attained for particular values of \( \kappa \), and not rely on extreme values of \( \tau \).

We consider three cases:

**Fixed Censoring** Without loss of generality we can assume the censoring point is constant at 0, as was considered in Honoré (1992) and Honoré and Powell (1994). To attain point identification, we need to find values of \( \kappa \) such that conditional on \( x_i, x_j \) we have:

\[ P(e_{ij}^0(\beta_0, \kappa_s, \kappa_t)|x_i, x_j) \geq e_{ji}^1(\beta_0, \kappa_s, \kappa_t)|x_i, x_j) = \frac{1}{2} \] (2.42)

That is, we need the moment points enable to attain a moment equality at \( \beta = \beta_0 \).

We can see that this can be accomplished at \( \kappa_s = -x_i' \beta_0, \kappa_t = -x_j' \beta_0 \). To see this, note
that
\[
e_{ij}^0(\beta_0, \kappa_s, \kappa_t) = \max(\epsilon_i, -\infty - x_i' \beta_0, -x_i' \beta_0, -x_j' \beta_0) = \max(\epsilon_i, -x_i' \beta_0, -x_j' \beta_0)
\]
(2.43)

and
\[
e_{ji}^1(\beta_0, \kappa_s, \kappa_t) = \max(\epsilon_j, -x_j' \beta_0, -x_i' \beta_0, -x_j' \beta_0) = \max(\epsilon_j, -x_i' \beta_0, -x_j' \beta_0)
\]
(2.45)

so the two are the same. So \(e_{ij}^0, e_{ji}^1\) are now identically distributed conditional on \(x_i, x_j\), and hence:
\[
P(\max(\epsilon_i, -x_i' \beta_0, -x_j' \beta_0) \geq \max(\epsilon_j, -x_j' \beta_0, -x_i' \beta_0) \mid x_i, x_j) = \frac{1}{2}
\]
(2.47)

which is the moment equality in Honoré and Powell (1994) which was designed for the special case of constant censoring or always observed censoring.

**Varying, Conditionally Independent Censoring** Our approach also applies to the random censoring case, where as in the Roy model, competing risk setting, the censoring variable is only partially observed. We note that neither Honoré (1992) nor Honoré and Powell (1994) can deal with setting. Our choices of \(\kappa\) will also enable the censoring variables to depend on observable covariates, as long as they are conditionally independent of disturbance terms. This is in contrast to and hence generalizes Honoré, Khan, and Powell (2002). In this setting the values of \(\kappa_i, \kappa_j\) which attain point identification are: \(\kappa_i = c_i - x_i' \beta_0\) and \(\kappa_j = c_j - x_j' \beta_0\). This will result in
\[
P(\max(\epsilon_i, c_i - x_i' \beta_0, c_j - x_j' \beta_0) \geq \max(\epsilon_j, c_j - x_j' \beta_0, c_i - x_i' \beta_0) \mid x_i, x_j) = \frac{1}{2}
\]
(2.48)

It is important to note we are not assuming the observation of \(c_i\) or \(c_j\) - the basic premise is that with \(\kappa_i, \kappa_j\) varying enough our set attained can reduce to a point.

**Dependent Censoring** Our framework can also allow for censoring variables that depend on unobservables in the latent equations. For this we could impose support conditions analogous to Khan and Tamer (2009), which for the problem at hand would be that there exists a subset of \(X \times X\) such that for \(x_i, x_j\) in this subset we have:
\[
P(c_i \leq x_i' \beta_0, c_j \leq x_j' \beta_0 \mid x_i, x_j) = 1
\]
(2.49)

Then, by \(\kappa_i, \kappa_j\) as set in the previous example we would again attain point identification.
Thus as we can see our frame can accommodate very general structures that are compatible with many Roy model and competing risk settings.

3 Preliminary Simulation Results

In this section we explore the finite sample properties of our proposed procedures by ways of a small scale monte carlo study. Results at this stage are preliminary as we only consider simulate designs for the cross sectional model, and report results only for the pairwise comparison procedure, with and without the additional censoring points.

Our simulation results are based on the latent equation:

\[ y^*_i = \alpha_0 + \beta_0 x_i + \epsilon_i \]  

where \( \alpha_0 = \beta_0 = 1 \), and \( x_i, \epsilon_i \) were each distributed standard normal and independent of each other. We considered two censoring designs, one with fixed censoring at 0, and the other where \( c_i \) was distributed standard normal, independently of \( (x_i, \epsilon_i) \). Table 1 reports the bias and RMSE for the two estimators at sample sizes of 50, 100, 200 for 401 replications.

To implement the estimators, for the first one, without artificial censoring, we set \( \tau \) to take values on an evenly spaced grid taking values between -2 and 2 with intervals of 0.1. For the artificially censoring moment points we chose one value of \( \tau \) at 0, and chose the artificial censoring points on an evenly spaced grid again between -2 and 2 with with intervals of 0.05. The estimators were computed by a grid search.

As the results indicate, both procedures seem to perform quite well, with RMSE's declining with the sample size, though the bias, while small, does not decline much with the sample size. Not surprisingly, the RMSE is noticeably larger for the randomly censored design than the fixed censoring design with both estimators. Overall, while the results are quite favorable, they can be sensitive to both the choice of number and location of the "moment points". This indicates that more theoretical work needs to be done to suggest finite sample choice of these variable.

| TABLE 1 |
| Simulation Results for MI Estimators |

16
<table>
<thead>
<tr>
<th></th>
<th>Fixed</th>
<th></th>
<th>Random</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Bias</td>
<td>RMSE</td>
<td>Mean Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>50 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI1</td>
<td>0.0672</td>
<td>0.2195</td>
<td>0.0529</td>
<td>0.2574</td>
</tr>
<tr>
<td>MI2</td>
<td>0.0456</td>
<td>0.3256</td>
<td>0.0677</td>
<td>0.2242</td>
</tr>
<tr>
<td>100 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI1</td>
<td>0.0696</td>
<td>0.1820</td>
<td>0.0784</td>
<td>0.2070</td>
</tr>
<tr>
<td>MI2</td>
<td>-0.1189</td>
<td>0.1791</td>
<td>0.071</td>
<td>0.2196</td>
</tr>
<tr>
<td>200 obs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI1</td>
<td>0.0096</td>
<td>0.1553</td>
<td>0.1101</td>
<td>0.1915</td>
</tr>
<tr>
<td>MI2</td>
<td>0.0429</td>
<td>0.1456</td>
<td>0.0544</td>
<td>0.1757</td>
</tr>
</tbody>
</table>

4 Conclusions

This paper considered identification and inference in a class of Roy models in cross section and panel data settings. 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), Honoré and Powell (1994).

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. In the former case, existing results for set inference Chernozhukov, Hong, and Tamer (2007), Kim (2008) can be applied.

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 consideration of a dynamic model, where lagged 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


