TRIMMED LAD AND LEAST SQUARES ESTIMATION OF TRUNCATED AND CENSORED REGRESSION MODELS WITH FIXED EFFECTS

By Bo E. Honoré

This paper considers estimation of truncated and censored regression models with fixed effects in panel data. Up until now, no estimator has been shown to be consistent as the cross section dimension increases with the time dimension fixed. Trimmed least absolute deviations (LAD) and trimmed least squares estimators are proposed for the case where the panel is of length two, and it is proven that they are consistent and asymptotically normal under suitable regularity conditions. It is not necessary to maintain parametric assumptions on the error terms to obtain this result. Because three of the four estimators are defined as minimizers of nondifferentiable functions, traditional methods cannot be used to establish asymptotic normality. Instead, the approach of Pakes and Pollard (1989) is used. A small scale Monte Carlo study demonstrates that these estimators can perform well in small samples. Despite their nonlinear nature, the estimators are easy to calculate in practice, as are consistent estimators of their asymptotic variances. Generalization of the estimators to panels of arbitrary length is briefly discussed.

KEYWORDS: Panel data, fixed effects, truncated and censored regression.

1. INTRODUCTION

Fixed effects play an important role in the analysis of panel data. If the econometric specification has the fixed effects entering linearly, then the remaining parameters of the model can be estimated by differencing the estimating equations. However, fixed effects can also be of importance in nonlinear models. For example, Heckman and MacCurdy's (1980) labor supply model implies that the decision of whether or not to work can be modelled as a discrete choice model with fixed effects, and that the actual labor supply will be a limited dependent variable with fixed effects. It is therefore of interest to study estimators of these models. As panel data are characterized by having many individuals observed over few time periods, it is most useful to consider the asymptotic properties of these estimators as the number of individuals increases with the number of observations per individual fixed. Estimation schemes that require estimation of the fixed effects will, in general, be inconsistent in this sense.

1 This research was supported by NSF Grants No. SES-8809352 and SES-9009879. Steve Cameron, Jeff Campbell, Tom Downes, Debbie Garvey, Jim Heckman, Charles Manski, Bruce Meyer, Whitney Newey, Rob Porter, Jim Powell, Richard Spady, two anonymous referees and the co-editor, Lars Hansen, made valuable comments on this paper, as did workshop participants at the University of Chicago, Northwestern University, University of Wisconsin, University of Minnesota, Indiana University, Carnegie Mellon University, Princeton University, New York University, Queens University, Cornell University, Harvard University, and Massachusetts Institute of Technology.

2 Throughout the paper, it is assumed that the parameter of interest is the vector of "remaining parameters." The fixed effects are treated as nuisance parameters.

533
Estimation of discrete choice models with fixed effects has been considered in the literature. If the underlying errors are independent with logistic distribution, then a conditional maximum likelihood estimator will be consistent and asymptotically normal (see Chamberlain (1984) for a discussion of this approach and for additional references). Manski (1987) has proposed a conditional maximum score estimator for the same model. The advantage of Manski’s estimator is that it is consistent under assumptions that are much weaker than those needed for the conditional maximum likelihood estimator.

This paper proposes estimators for limited dependent variable models with fixed effects. These models are usually estimated by maximizing a likelihood function over all the parameters, including the fixed effects. As mentioned, these estimators will generally not have the desired asymptotic properties. In contrast, the estimators proposed here are consistent and asymptotically normal as the number of individuals approaches infinity with the number of observations per individual fixed. The idea behind the estimators can be thought of as a bivariate generalization of the idea behind Powell’s (1986) trimmed least squares estimators for Tobit models (without fixed effects). Like Powell’s estimators, the estimators presented here are semiparametric. It is not necessary to assume a parametric form for the disturbances. Nor is it necessary to assume homoskedasticity across individuals.

The problems with estimating limited dependent variable models with fixed effects from panel data are most acute when the time dimension of the panel is low. For the greater part of this paper, we therefore restrict attention to the case with two time periods. The notation for this case is also simpler. In Section 7, we briefly discuss extensions of the estimators to more than two periods.

Section 2 of this paper defines the estimators and explains why they might work. Consistency and asymptotic normality of the estimators are investigated in Sections 3 and 4, respectively. The asymptotic variances derived in Section 4 will, in practice, have to be estimated. In Section 5, we suggest consistent estimators of these variances. Section 6 presents the results of a small scale Monte Carlo experiment, which illustrates that the large sample properties of the estimators can provide good approximations in small samples and that it is feasible computationally to calculate the estimators. The latter is of interest because it can be cumbersome to estimate a large number of fixed effects. Extensions of the estimators to the case where there are more than two observations for each individual, are discussed in Section 7, which concludes the paper. All proofs are in the appendices.

2. THE ESTIMATORS

It is useful to think about the data being generated as transformations of unobserved latent variables $Y_1^*$ and $Y_2^*$ given by

\begin{equation}
Y_t^* = \alpha + X_t \beta + \epsilon_t \quad \text{for} \quad t = 1, 2,
\end{equation}

where $X_1$ and $X_2$ are $K$-dimensional (row-)vectors of explanatory variables, $\beta$ is
TRIMMED LAD ESTIMATION

the parameter (column-)vector of interest, and \( \alpha \) is the fixed effect. \( \varepsilon_1 \) and \( \varepsilon_2 \) are the error terms.

Two sampling schemes are considered. For the truncated regression model, the observations \((Y_{1i}, X_{1i}, Y_{2i}, X_{2i}): i = 1, \ldots, n\) are a random sample from the distribution of \((Y^*_1, X_1, Y^*_2, X_2): i = 1, \ldots, n\) induced by conditioning on the event \(Y^*_1 > 0 \) and \(Y^*_2 > 0\). For the censored regression model, we observe \(\{Y_{it}, X_{it}\}: t = 1, 2, i = 1, \ldots, n\) where \(Y_{it} = \max\{0, Y^*_i\}\), and \(Y^*_i \) and \(X_{it}\) are distributed as in (2.1). In terms of Heckman and MaCurdy’s (1980) labor supply model, the truncated model would be relevant if the sample consists of continuously working individuals, whereas the censored model is appropriate if the sample consists of a random sample of individuals, with hours worked reported as 0 if the individual does not work. From a practical point of view, the latter is the more interesting model, but since the truncated sampling is more extreme, it is also of interest to consider this model. The notation will be simplified if we define \(A_X = X_1 - X_2\) and \(A_Y = Y_1 - Y_2\).

To facilitate the exposition, this is stated as a proposition.

**Proposition 1:** If \(\varepsilon_1\) and \(\varepsilon_2\) in (2.1) are independent and identically distributed conditional on \((X_1, X_2, \alpha)\), then the distribution of \((Y^*_1, Y^*_2)\) conditional on \((X_1, X_2)\) is symmetric around the 45°-line through \((\Delta X \beta, 0)\).

The purpose of this section is to point out that the symmetry in Proposition 1 suggests orthogonality conditions that must hold at the true parameter values. These in turn suggest objective functions that can be used to estimate \(\beta\) in (2.1) when the data are either censored or truncated, as described above. The idea of using the symmetry of \((Y^*_1, Y^*_2)\) around the 45°-line through \((\Delta X \beta, 0)\) can be thought of as a two-dimensional extension of Powell’s (1986) idea of using symmetry of \(Y\) around \(X \beta\) to construct an estimator of \(\beta\) in an ordinary Tobit model. In Powell’s work it is assumed that the distribution of the errors is symmetric. This is not the case here. The symmetry is a consequence of the assumption that \(\varepsilon_1\) and \(\varepsilon_2\) are independent and identically distributed.

It is easiest to illustrate the ideas graphically. Let \((x_1, x_2)\) be arbitrary. The distribution of \((Y^*_1, Y^*_2)\) conditional on \((X_1, X_2) = (x_1, x_2)\) is then symmetric around the line \(LL'\) in Figures 1A and 1B (in Figure 1A, it is assumed that \(\Delta x \beta \geq 0\), where Figure 1B illustrates the situation \(\Delta x \beta < 0\)).

Consider first the truncated case. The symmetry of the distribution of \((Y^*_1, Y^*_2)\) around \(LL'\) means that the probability that \((Y^*_1, Y^*_2)\) falls in the region \(A_1\) equals the probability that it falls in the region \(B_1\) (the sets \(A_1, A_2, B_1,\) and \(B_2\) are all defined in Figures 1A and 1B). Observations in the regions \(A_1\) and \(B_1\)
are not affected by the truncation, so the same statement is true for the observed variables, \((Y_1, Y_2)\). Therefore

\[
E[(I\{(Y_1, Y_2) \in A_1\} - 1\{(Y_1, Y_2) \in B_1\}) \Delta X] = 0,
\]

where \(I\{\cdot\}\) represents the indicator function.

Using the same argument as above, the expected vertical distance from \((Y_1^*, Y_2^*)\) in \(A_1\) to the line \(LL'\) equals the expected horizontal distance from \((Y_1^*, Y_2^*)\) in \(B_1\) to the line \(LL'\), where both expectations are conditional on \((X_1, X_2)\). For \((Y_1^*, Y_2^*)\) in \(A_1\), the vertical distance to \(LL'\) can be written as \(- (Y_1^* - Y_2^* - \Delta x \beta)\) whereas for \((Y_1^*, Y_2^*)\) in \(B_1\) the horizontal distance to \(LL'\)

\[3\] The moment condition in (2.2) is also the basis for the estimator of the truncated regression model (without fixed effects) proposed by Bhattacharya, Chernoff, and Yang (1983).
is $Y_1^* - Y_2^* - \Delta x \beta$. Therefore

$$\begin{align*}
(2.3) \quad & E\left[1\{(Y_1, Y_2) \in A_1 \cup B_1\}(Y_1 - Y_2 - \Delta X \beta) \Delta X\right] \\
& = E\left[E\left[1\{(Y_1, Y_2) \in A_1\}(Y_1 - Y_2 - \Delta X \beta)\mid X_1, X_2\right]\right] \\
& \quad + E\left[1\{(Y_1, Y_2) \in B_1\}(Y_1 - Y_2 - \Delta X \beta)\mid X_1, X_2\right]) \Delta X
\end{align*}$$

= 0.

The two estimators for the truncated regression model, $\hat{\beta}_1$ and $\hat{\beta}_2$, are defined by minimization of objective functions that have as first order conditions that the sample analogs of (2.2) and (2.3), respectively, are satisfied. Specifically, $\hat{\beta}_1$ and $\hat{\beta}_2$ are defined by minimization of $Q_n(b)$ and $R_n(b)$, where

$$Q_n(b) = \sum_{i=1}^n |\Delta Y_i - \Delta X_i b| \mathbb{1}\{Y_{i1} \geq \Delta X_i b, Y_{i2} \geq -\Delta X_i b\}$$

$$+ |Y_{i1}| \mathbb{1}\{Y_{i1} \geq \Delta X_i b, Y_{i2} < -\Delta X_i b\}$$

$$+ |Y_{i2}| \mathbb{1}\{Y_{i1} < \Delta X_i b, Y_{i2} \geq -\Delta X_i b\}$$

$$= \sum_{i=1}^n \psi(Y_{i1}, Y_{i2}, \Delta X_i b)$$

and

$$R_n(b) = \sum_{i=1}^n (\Delta Y_i - \Delta X_i b)^2 \mathbb{1}\{Y_{i1} \geq \Delta X_i b, Y_{i2} \geq -\Delta X_i b\}$$

$$+ Y_{i1}^2 \mathbb{1}\{Y_{i1} \geq \Delta X_i b, Y_{i2} < -\Delta X_i b\}$$

$$+ Y_{i2}^2 \mathbb{1}\{Y_{i1} < \Delta X_i b, Y_{i2} \geq -\Delta X_i b\}$$

$$= \sum_{i=1}^n \psi(Y_{i1}, Y_{i2}, \Delta X_i b)^2,$$

where the function $\psi(z_1, z_2, \delta)$ is defined for $z_1 > 0$ and $z_2 > 0$ by

$$\psi(z_1, z_2, \delta) = \begin{cases} 
z_1, & \text{for } \delta \leq -z_2; \\
|z_1 - z_2 - \delta|, & \text{for } -z_2 < \delta < z_1; \\
z_2, & \text{for } z_1 \leq \delta.
\end{cases}$$

Now turn to the censored regression model. By the same argument that led to (2.2), the probability of $(Y_1^*, Y_2^*)$ falling in $A = A_1 \cup A_2$ equals the probability that it will fall in $B = B_1 \cup B_2$ (where both probabilities are conditional on $(X_1, X_2)$). As neither of these probabilities is affected by censoring, the same is true in the censored sample. This implies

$$\begin{align*}
(2.4) \quad & E\left[\left(1\{(Y_1, Y_2) \in A_1 \cup A_2\} - 1\{(Y_1, Y_2) \in B_1 \cup B_2\}\right) \Delta X\right] = 0.
\end{align*}$$

As in the truncated model, the expected vertical distance from a $(Y_1^*, Y_2^*)$ in $A$ to the boundary of $A$, equals the expected horizontal distance from a
\[ (Y_1^*, Y_2^*) \text{ in } B \text{ to the boundary of } B \text{ (again these expectations are conditional on } (X_1, X_2)). \text{ As neither of these distances is affected by censoring, the same is true for the distribution of the censored observations } (Y_1, Y_2). \text{ Therefore} \\
\]

\[
(2.5) \quad E \left[ \left( 1 \{ (Y_1, Y_2) \in A_1 \} (Y_1 - Y_2 - AX \beta) - 1 \{ (Y_1, Y_2) \in A_2 \} \right) \times \left( Y_2 - \max \{ 0, -AX \beta \} \right) + 1 \{ (Y_1, Y_2) \in B_1 \} (Y_1 - Y_2 - AX \beta) \right] = 0.
\]

The proposed estimators for the censored regression model, \( \hat{\beta}_3 \) and \( \hat{\beta}_4 \), are defined by minimization of objective functions that have as first order conditions that the sample analogs of (2.4) and (2.5), respectively, are satisfied. \( \hat{\beta}_3 \) and \( \hat{\beta}_4 \) are defined by minimization of \( S_n(b) \) and \( T_n(b) \), where

\[
S_n(b) = \sum_{i=1}^{n} (1 - 1 \{ Y_{i1} \leq AX_i b, Y_{i2} \leq 0 \}) \times \left( 1 - 1 \{ Y_{i2} \leq -AX_i b, Y_{i1} \leq 0 \} \right) |Y_{i1} - Y_{i2} - AX_i b| = \sum_{i=1}^{n} \varphi(Y_{i1}, Y_{i2}, AX_i b),
\]

where

\[
\varphi(z_1, z_2, \delta) = \begin{cases} 
0, & \text{for } z_1 \leq \max \{ 0, \delta \} \text{ and } z_2 \leq \max \{ 0, -\delta \}, \\
|z_1 - z_2 - \delta|, & \text{otherwise},
\end{cases}
\]

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{psi_plot.png}
\caption{Plot of \( \psi \).}
\end{figure}
and

\[ T_n(b) = \sum_{i=1}^{n} \left( \max \{ Y_{i1}, \Delta X_i b \} - \max \{ Y_{i2}, -\Delta X_i b \} - \Delta X_i b \right)^2 \\
+ 2 \times 1 \{ Y_{i1} < \Delta X_i b \} (\Delta X_i b - Y_{i1}) Y_{i2} \\
+ 2 \times 1 \{ Y_{i2} < -\Delta X_i b \} (-\Delta X_i b - Y_{i2}) Y_{i1} \\
= \sum_{i=1}^{n} \chi(Y_{i1}, Y_{i2}, \Delta X_i b), \]

where

\[ \chi(z_1, z_2, \delta) = \begin{cases} 
z_1^2 + 2z_1(-z_2 - \delta), & \text{for } \delta \leq -z_2, \\
(z_1 - z_2 - \delta)^2, & \text{for } -z_2 < \delta < z_1, \\
z_2^2 + 2z_2(\delta - z_1), & \text{for } z_1 \leq \delta.
\]

Figures 2–5 graph the functions \( \psi, \psi^2, \varphi, \) and \( \chi. \)

Note that \( S_n \) is piecewise linear and convex. \( T_n \) is continuously differentiable and convex and twice differentiable except at a finite number of points. This implies that \( \hat{\beta}_3 \) and especially \( \hat{\beta}_4 \) are very easy to calculate. The objective functions for the truncated case, \( Q_n \) and \( R_n, \) are not convex, and it is clear from Figures 2 and 3 that they are likely to have numerous local minima.

3. CONSISTENCY

In this section, we study the consistency of the estimators defined in the previous section. First, the two models considered in this paper are formally stated.
Assumption S.1 (Truncated Sampling): The data are generated as i.i.d. observations from (2.1) conditional on $Y_1 > 0$ and $Y_2 > 0$, where it is assumed that the unconditional probability that $Y_1 > 0$ and $Y_2 > 0$ exceeds zero.

Assumption S.2 (Censored Sampling): The data are generated as i.i.d. observations from (2.1), except that $Y_1 = \max\{0, Y_1^*\}$ and $Y_2 = \max\{0, Y_2^*\}$ are observed rather than $Y_1^*$ and $Y_2^*$. It is assumed that $P(Y_1 > 0, Y_2 > 0) > 0$.

In addition to defining the models, Assumptions S.1 and S.2 also require random sampling. This assumption is made for convenience and is not essential.
It is possible to derive the asymptotic properties of the estimators considered here if the observations are i.n.i.d., or have some dependence.$^4$

To prove consistency of $\hat{\beta}_1$ and $\hat{\beta}_2$, we state conditions under which the assumptions of Theorem 4.1.1 in Amemiya (1985) are satisfied. These are assumptions of compactness of the parameter space, uniform convergence of the objective function, and identification (i.e. the limit of the objective function is uniquely minimized at the true parameter value, $\beta$). Newey and Powell (1987) have demonstrated that if an estimator is defined by minimization of a convex function, then the assumption that the parameter space is compact can be.

$^4$Pakes and Pollard's (1989) result cannot be used to prove asymptotic normality without the i.i.d. assumption. However, generalizations of Huber's (1965) papers (see Powell (1984) and Weiss (1991) or the results in Andrews (1991a, b)) could be used.
eliminated. The consistency of $\hat{\beta}_3$ and $\hat{\beta}_4$ is therefore proven by satisfying the conditions of that paper.

For the truncated case, compactness of the parameter space is assumed and, in anticipation of the discussion of asymptotic normality, it is also assumed that $\beta$ is in the interior of the parameter space:

**Assumption P.1** (Compact Parameter Space): The parameter space, $B$, is compact, and the true value of the parameter is an interior point of $B$, $\beta \in \text{int } B$.

Uniform convergence of the objective function obtains under the right uniform law of large numbers. We will make use of the following assumptions.

**Assumption M.1** (Moment Condition): $E[X_1]$ and $E[X_2]$ are finite.

**Assumption M.2** (Moment Condition): $E[\|X_1\|^2]$, $E[\|X_2\|^2]$, $E[\alpha \|\Delta X\|]$, $E[\epsilon_1 \|\Delta X\|]$, and $E[\epsilon_2 \|\Delta X\|]$ are all finite.

LAD-type estimators are, in general, not identified unless it is assumed that the observations are continuously distributed. For $\hat{\beta}_1$ and $\hat{\beta}_3$ we therefore make the following assumption.

**Assumption E.1**: The conditional distribution of $(\epsilon_1, \epsilon_2)$ given $(\alpha, X_1, X_2)$ is continuous with finite density.

To prove identification for the truncated model, the following assumption is made.

**Assumption E.2** (Symmetry and Unimodality): The distribution of $\epsilon_1 - \epsilon_2$ conditional on $\epsilon_1 + \epsilon_2$ and on $(\alpha, X_1, X_2)$ is strictly unimodal$^5$ and symmetric around 0.

For the unimodality in Assumption E.2 to be well-defined, Assumption E.1 will be maintained whenever Assumption E.2 is made. The unimodality assumption in E.2 plays the same role as the unimodality assumption needed for Powell's symmetrically trimmed least squares estimator for the truncated regression model.

For the censored model, unimodality is not needed. We make the following assumption.$^6$

**Assumption E.3** (Symmetry): The distribution of $\epsilon_1 - \epsilon_2$ conditional on $\epsilon_1 + \epsilon_2$ and on $(\alpha, X_1, X_2)$, is symmetric around 0.

$^5$ We say that a continuous distribution is strictly unimodal, if its density has a unique maximum, and if it is monotone on either side of the maximum with strict monotonicity in a neighborhood around the maximum. This is the definition used by Powell (1986).

$^6$ Assumption E.3 can be restated as an assumption that $\epsilon_1$ and $\epsilon_2$ are exchangeable.
More readily interpreted conditions implying E.2 and E.3, are given later in this section. We also need to make an assumption about the rank of the regressors:

**Assumption R.1 (Full Rank of Regressors):** There is no proper linear subspace of $R^K$ containing the random variable $1\{P(Y_1 > 0, Y_2 > 0|X_1, X_2) > 0\} \Delta X$ with probability 1.

With these assumptions, the consistency result can be stated.

**Theorem 1 (Consistency):** The estimators $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ are consistent under the following conditions:

(i) $\hat{\beta}_1$ defined by minimization of $Q_n(b)$ is strongly consistent under Assumptions P.1, S.1, M.1, E.1, E.2, and R.1.

(ii) $\hat{\beta}_2$ defined by minimization of $R_n(b)$ is strongly consistent under Assumptions P.1, S.1, M.2, E.1, E.2, and R.1.

(iii) $\hat{\beta}_3$ defined by minimization of $S_n(b)$ is strongly consistent under Assumptions S.2, M.1, E.1, E.3, and R.1.

(iv) $\hat{\beta}_4$ defined by minimization of $T_n(b)$ is strongly consistent under Assumptions S.2, M.2, E.3, and R.1.

We now return to the interpretation of Assumptions E.2 and E.3. It is not clear, a priori, whether these assumptions are reasonable. The following lemma states conditions that imply that Assumptions E.2 and E.3 are satisfied.

**Lemma 1:** If, conditional on $(\alpha, X_1, X_2)$, $\varepsilon_1$ and $\varepsilon_2$ are independent, and identically and continuously distributed, then Assumption E.3 is satisfied. If, in addition, the (common) marginal density of $\varepsilon_1$ and $\varepsilon_2$ is strictly log-concave (i.e., the logarithm of the density is strictly concave), then Assumption E.2 is also satisfied.

The assumption that $\varepsilon_1$ and $\varepsilon_2$ are independent in Lemma 1 is less restrictive than it may seem. The fixed effect can capture some dependence between the error terms. For example, the assumptions of Lemma 1 are satisfied if the error terms are jointly normal, with equal variance but arbitrary positive correlation. Dependent normals with positive correlation can be written as a (normal) fixed effect plus independent normals—the dependence can be captured in the fixed effect.

The assumptions of Lemma 1 allow for heteroskedasticity across individuals, but homoskedasticity is assumed over time (for a given individual). This is not just a technical assumption. Without it, the estimators in this paper would be inconsistent.
4. ASYMPTOTIC NORMALITY

The objective functions $Q_n$, $R_n$, $S_n$, and $T_n$ are not twice differentiable. This means that standard results (see, for example, Amemiya (1985)) cannot be used to prove asymptotic normality of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$. Powell (1984, 1986) proves asymptotic normality for closely related estimators by verifying the conditions of Huber (1965). In this paper, the approach of Pakes and Pollard (1989) is adopted. This choice is made for convenience—it certainly seems possible to verify the conditions of Huber (1965) for the estimators defined in this paper, but it is easier to verify the conditions in Pakes and Pollard (1989). Alternatively, the results of Pollard (1985) or Andrews (1991a, b) could have been used.

In order to prove asymptotic normality of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, and $\hat{\beta}_4$ we make the following definitions:

\[
V_1 = E\{1\{-Y_2 < \Delta X\beta < Y_1\}\Delta X'\Delta X\},
\]
\[
V_2 = E\{1\{-Y_2 < \Delta X\beta < Y_1\}(\Delta Y - \Delta X\beta)^2 \Delta X'\Delta X\},
\]
\[
V_3 = E\{1\{Y_1 < Y_2 + \Delta X\beta, Y_2 > \max \{0, -\Delta X\beta\}\}
\cup \{Y_1 > Y_2 + \Delta X\beta, Y_1 > \max \{0, \Delta X\beta\}\}\Delta X'\Delta X\},
\]
\[
V_4 = E\{(Y_2^21\{Y_1 < \Delta X\beta\} + Y_1^21\{\Delta X\beta < -Y_2\}
+(Y_1 - Y_2 - \Delta X\beta)^21\{-Y_2 < \Delta X\beta < Y_1\}\}\Delta X'\Delta X\},
\]
\[
\Gamma_1 = E\left\{\left(2f_{Y_1-Y_3}(\Delta X\beta|X_1, X_2, \alpha) - 1\{\Delta X\beta \geq 0\}f_{Y_1}(\Delta X\beta|X_1, X_2, \alpha) - 1\{\Delta X\beta < 0\}f_{Y_2}(-\Delta X\beta|X_1, X_2, \alpha)\right)\Delta X'\Delta X\right\},
\]
\[
\Gamma_2 = E\left\{1\{\Delta X\beta \geq 0\}f_{Y_1}(\Delta X\beta|X_1, X_2, \alpha)E[Y_2|Y_1 = \Delta X\beta, X_1, X_2, \alpha]
+1\{\Delta X\beta < 0\}f_{Y_2}(-\Delta X\beta|X_1, X_2, \alpha)
\times E[Y_1|Y_2 = -\Delta X\beta, X_1, X_2, \alpha]
-P(Y_1 > \Delta X\beta, Y_2 > -\Delta X\beta|X_1, X_2, \alpha)\Delta X'\Delta X\right\},
\]
\[
\Gamma_3 = E\left\{2f_{Y_1-Y_3}(\Delta X\beta|X_1, X_2, \alpha, Y_1 > 0, Y_2 > 0)
+1\{\Delta X\beta \geq 0\}P(Y_2 \leq 0|\alpha, X_1, X_2)
\times f_{Y_1}(\Delta X\beta|X_1, X_2, \alpha, Y_2 \leq 0)
+1\{\Delta X\beta < 0\}P(Y_1 \leq 0|\alpha, X_1, X_2)
\times f_{Y_2}(-\Delta X\beta|X_1, X_2, \alpha, Y_1 \leq 0)\Delta X'\Delta X\right\},
\]
\[
\Gamma_4 = E\{1\{-Y_2 < \Delta X\beta < Y_1\}\Delta X'\Delta X\}.
\]

The following assumptions will be used to prove asymptotic normality of the estimators.
ASSUMPTION V (Rank of Covariance Matrices): (1) The expectations in \( V_1 \) and \( \Gamma_1 \) are finite and \( V_1 \) and \( \Gamma_1 \) are of full rank. (2) The expectations in \( V_2 \) and \( \Gamma_2 \) are finite and \( V_2 \) and \( \Gamma_2 \) are of full rank. (3) The expectations in \( V_3 \) and \( \Gamma_3 \) are finite and \( V_3 \) and \( \Gamma_3 \) are of full rank. (4) The expectations in \( V_4 \) and \( \Gamma_4 \) are finite and \( V_4 \) and \( \Gamma_4 \) are of full rank.

ASSUMPTION M.3 (Moment Condition): \( E[\|X_1\|^4], E[\|X_2\|^4], E[\alpha^2\|\Delta X\|^2], E[\epsilon_1^2\|\Delta X\|^2] \) and \( E[\epsilon_2^2\|\Delta X\|^2] \) are all finite.

ASSUMPTION E.4: There exists a number \( M < \infty \) such that \( f_{\epsilon_1 - \epsilon_2}(\cdot|X_1, X_2, \alpha) < M \), \( f_{\epsilon_1}(\cdot|X_1, X_2, \alpha) < M \), and \( f_{\epsilon_2}(\cdot|X_1, X_2, \alpha) < M \) for almost all \((X_1, X_2, \alpha)\).

ASSUMPTION E.5: There exists an \( M < \infty \) such that \( f_{\epsilon_1}(\cdot, \epsilon_2, X_1, X_2, \alpha) < M \) for almost all \((\epsilon_2, X_1, X_2, \alpha)\) and \( f_{\epsilon_2}(\cdot, \epsilon_1, X_1, X_2, \alpha) < M \) for almost all \((\epsilon_1, X_1, X_2, \alpha)\).

With these assumptions, we can derive the asymptotic distributions of the estimators defined in Section 2.

**Theorem 2 (Asymptotic Normality):**

(i) \( \hat{\beta}_1 \) is asymptotically normal under Assumptions P.1, S.1, M.2, E.1, E.2, R.1, V.1, and E.4:

\[
\sqrt{n} (\hat{\beta}_1 - \beta) \rightarrow N(0, \Gamma_1^{-1}V_1 \Gamma_1^{-1}).
\]

(ii) \( \hat{\beta}_2 \) is asymptotically normal under Assumptions P.1, S.1, M.3, E.1, E.2, R.1, V.2, and E.5:

\[
\sqrt{n} (\hat{\beta}_2 - \beta) \rightarrow N(0, \Gamma_2^{-1}V_2 \Gamma_2^{-1}).
\]

(iii) \( \hat{\beta}_3 \) is asymptotically normal under Assumptions S.2, M.2, E.1, E.3, R.1, V.3, and E.4:

\[
\sqrt{n} (\hat{\beta}_3 - \beta) \rightarrow N(0, \Gamma_3^{-1}V_3 \Gamma_3^{-1}).
\]

(iv) \( \hat{\beta}_4 \) is asymptotically normal under Assumptions S.2, M.3, E.1, E.3, V.4, and R.1:

\[
\sqrt{n} (\hat{\beta}_4 - \beta) \rightarrow N(0, \Gamma_4^{-1}V_4 \Gamma_4^{-1}).
\]
Assumptions E.4 and E.5 guarantee that the expectations of the objective functions (evaluated at $\beta$) are twice differentiable and that the second derivative can be found by interchanging differentiation and expectation. The latter is not necessary and Assumptions E.4 and E.5 could be replaced by alternative assumptions that will make the expected objective functions twice differentiable.

5. ESTIMATING THE ASYMPTOTIC VARIANCES

For the asymptotic distributions in Theorem 2 to be useful in practice, we must be able to consistently estimate $V_1$, $V_2$, $V_3$, $V_4$, $\Gamma_1$, $\Gamma_2$, $\Gamma_3$, and $\Gamma_4$. This section presents one set of consistent estimators for these matrices.

The estimators for $V_1$, $V_2$, $V_3$, $V_4$, and $\Gamma_4$ are straightforward, whereas the estimators for $\Gamma_1$, $\Gamma_2$, and $\Gamma_3$ are more complicated because they involve estimation of densities.

$V_1$, $V_2$, $V_3$, $V_4$, and $\Gamma_4$ can be estimated by their sample analogs:

$$
\hat{V}_1 = \frac{1}{n} \sum_{i=1}^{n} 1\{-Y_{i2} < \Delta X_i \hat{\beta}_1 < Y_{i1}\} \Delta X'_i \Delta X_i,
$$

$$
\hat{V}_2 = \frac{1}{n} \sum_{i=1}^{n} 1\{-Y_{i2} < \Delta X_i \hat{\beta}_2 < Y_{i1}\}(\Delta Y_i - \Delta X_i \hat{\beta}_2)^2 \Delta X'_i \Delta X_i,
$$

$$
\hat{V}_3 = \frac{1}{n} \sum_{i=1}^{n} 1\{(\Delta Y_i < \Delta X_i \hat{\beta}_3, Y_{i2} > \max\{0, -\Delta X_i \hat{\beta}_3\}) \cup (\Delta Y_i > \Delta X_i \hat{\beta}_3, Y_{i1} > \max\{0, \Delta X_i \hat{\beta}_3\})\} \Delta X'_i \Delta X_i,
$$

$$
\hat{V}_4 = \frac{1}{n} \sum_{i=1}^{n} \left(Y_{i2} 1\{Y_{i1} \leq \Delta X_i \hat{\beta}_4\} + Y_{i1}^2 1\{\Delta X_i \hat{\beta}_4 \leq -Y_{i2}\} \right) + (Y_{i1} - Y_{i2} - \Delta X_i \beta)^2 1\{-Y_{i2} < \Delta X_i \beta < Y_{i1}\} \Delta X'_i \Delta X_i
$$

and

$$
\hat{\Gamma}_4 = \frac{1}{n} \sum_{i=1}^{n} 1\{-Y_{i2} < \Delta X_i \hat{\beta}_4 < Y_{i1}\} \Delta X'_i \Delta X_i.
$$

It is less obvious how to construct sample analogs of $\Gamma_1$, $\Gamma_2$, and $\Gamma_3$ because they involve the densities and conditional expectations. In general, these can be estimated by the method of kernels. Using uniform kernels, Powell (1984, 1986) constructed consistent estimators of the variances of estimators similar to those considered in this paper. It is possible (see Honoré (1989)) to use the same approach to form consistent estimators of $\Gamma_l$ ($l = 1, 2, 3$), but the proof of Theorem 2 suggests estimators that are simpler to analyze.

Consider, for example, $\Gamma_3$. From the proof of Theorem 2, it is clear that $\Gamma_3$ is constructed in such a way that it equals the derivative of the expected values of the “first order conditions” for minimization of $S_n$, evaluated at the true
parameter value. In other words, the \((j, k)\)th element of \(\Gamma_3\) is

\[
\Gamma_3^{(j, k)} = \frac{d}{db_k} E \left[ \left( 1 \{ Y_2 > 0, Y_2 > Y_1 - \Delta X b \} - 1 \{ Y_1 > 0, Y_1 > Y_2 + \Delta X b \} \right) \Delta X^{(j)} \right]
\]
evaluated at \(b = \beta\), where \(\Delta X^{(j)}\) is the \(j\)th coordinate of \(\Delta X\). One “natural” estimator of \(\Gamma_3\) is one that modifies this expression by replacing \(\beta\) by \(\hat{\beta}_3\), replacing the expectation by a sample average, and by replacing the differentiation by a finite difference which decreases with sample size. This suggests estimating \(\Gamma_3^{(j, k)}\) by

\[
\hat{\Gamma}_3^{(j, k)} = \frac{1}{\omega_n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( 1 \{ Y_{i2} > 0, Y_{i2} > Y_{i1} - \Delta X_i (\hat{\beta}_3 + \omega_n u_k) \} - 1 \{ Y_{i1} > 0, Y_{i1} > Y_{i2} + \Delta X_i (\hat{\beta}_3 + \omega_n u_k) \} \right) \Delta X_i^{(j)} \right.
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \left( 1 \{ Y_{i2} > 0, Y_{i2} > Y_{i1} - \Delta X_i \hat{\beta}_3 \} - 1 \{ Y_{i1} > 0, Y_{i1} > Y_{i2} + \Delta X \hat{\beta}_3 \} \right) \Delta X_i^{(j)}
\]

where \(u_k\) is a unit vector with 1 in its \(k\)th place and \(\omega_n\) decreases to 0 with sample size.

Estimators of \(\Gamma_1\) and \(\Gamma_2\) can be constructed in the same way:

\[
\hat{\Gamma}_1^{(j, k)} = \frac{1}{\omega_n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( -1 \{ -Y_{i2} < \Delta X_i (\hat{\beta}_1 + \omega_n u_k) < \Delta Y_i \} + 1 \{ \Delta Y_i < \Delta X_i (\hat{\beta}_1 + \omega_n u_k) < Y_{i1} \} \right) \Delta X_i^{(j)} \right.
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \left( -1 \{ -Y_{i2} < \Delta X_i \hat{\beta}_1 < \Delta Y_i \} + 1 \{ \Delta Y_i < \Delta X_i \hat{\beta}_1 < Y_{i1} \} \right) \Delta X_i^{(j)}
\]

and

\[
\hat{\Gamma}_2^{(j, k)} = \frac{1}{\omega_n} \left( \frac{1}{n} \sum_{i=1}^{n} \left( -1 \{ -Y_{i2} < \Delta X_i (\hat{\beta}_2 + \omega_n u_k) < Y_{i1} \} \times (Y_{i1} - Y_{i2} - \Delta X_i (\hat{\beta}_2 + \omega_n u_k)) \Delta X_i^{(j)} \right.
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} \left( -1 \{ -Y_{i2} < \Delta X_i \hat{\beta}_2 < Y_{i1} \} \times (Y_{i1} - Y_{i2} - \Delta X_i \hat{\beta}_2) \Delta X_i^{(j)} \right)
\]
If $\omega_n$ goes to zero too quickly, the term $1/\omega_n$ will prevent the estimators of $\Gamma_1, \Gamma_2,$ and $\Gamma_3$ from being consistent. The following assumption prevents this.

**Assumption W.1:** The sequences $\{\omega_n\}_{n=1}^\infty$ converges to 0 in probability and $n^{-1/2}\omega_n^{-1} = O_p(1)$.

**Theorem 3:** For $l=1, \ldots, 4$, $V_l$ and $\Gamma_l$ are consistent under Assumption W.1 and the assumptions made for $\beta_l$ in Theorem 2.

It is clear that $\omega_n$ need not be the same for different columns of $\Gamma_l$. Furthermore, $\omega_n$ can be arbitrarily data-dependent, as long as Assumption W.1 is satisfied.

6. Monte Carlo Investigation

In this section, we summarize a small scale Monte Carlo study designed to illustrate the small sample properties of the estimators defined in Section 2 and the computational feasibility of the estimators. Only $\hat{\beta}_4$ is considered. As noted earlier, the censored regression model is probably more interesting from an applied point of view than the truncated regression model. And since $\hat{\beta}_4$ is defined in terms of minimizing a differentiable function, it is simpler to calculate than $\hat{\beta}_3$. By considering only $\hat{\beta}_4$, we also avoid making arbitrary decisions about $\omega_n$, when studying the small sample behavior of the "$t$ statistics" implied by the large sample distributions.

All the results presented in this paper are based on 1000 replications from the model

\[ Y_{it} = \max \left\{ 0, \alpha_i + \sum_{k=1}^K X_{kit} \beta_k + \epsilon_{it} \right\} \quad \text{for} \quad t=1,2; \ i=1,\ldots,n. \]

The first three specifications take $K=2$ and $X_{1it} = \alpha_i + \eta_{it}$. The three specifications illustrate the behavior of the estimator for different sample sizes and differ only in their distributional assumptions. The results from these specifications are given in Table I. The table reports the true parameter values, the estimated bias and root mean squared error of $\hat{\beta}_4$, as well as the root mean squared error implied by the asymptotic distribution in Theorem 2. As the estimators may not have finite moments in any sample size, we also report the quartiles and the median absolute error (MAE) of the estimator, as well as the median absolute error predicted by the asymptotic distribution (AMAE).

In Specification 1, the random variables $\alpha_i$, $\eta_{i1}$, $\eta_{i2}$, $X_{2i1}$, $X_{2i2}$, $\epsilon_{i1}$, and $\epsilon_{i2}$ are independent and normally distributed with mean 0 and variance 1. The true value of $\beta$ is $(1,1)$. With this specification $P(Y_{i1} > 0, Y_{i2} > 0) = 0.35$, $P(Y_{i1} > 0,$

---

7 Honoré (1989) reports the results of similar experiments for $\hat{\beta}_1$ and $\hat{\beta}_3$.

8 All the calculations reported in this section were performed in GAUSS386. The pseudo-random generators in GAUSS were used to generate the samples.
### Table I

#### Results for Different Sample Sizes

<table>
<thead>
<tr>
<th>Specification</th>
<th>n = 20</th>
<th>n = 50</th>
<th>n = 200</th>
<th>n = 500</th>
<th>n = 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>b1</td>
<td>1.00</td>
<td>0.057</td>
<td>0.648</td>
<td>0.403</td>
<td>0.037</td>
</tr>
<tr>
<td>b2</td>
<td>1.00</td>
<td>0.078</td>
<td>0.764</td>
<td>0.398</td>
<td>0.072</td>
</tr>
<tr>
<td>Specification 2.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 20</td>
<td>1.00</td>
<td>0.055</td>
<td>0.648</td>
<td>0.403</td>
<td>0.037</td>
</tr>
<tr>
<td>b2</td>
<td>1.00</td>
<td>0.078</td>
<td>0.766</td>
<td>0.398</td>
<td>0.072</td>
</tr>
<tr>
<td>Specification 3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 20</td>
<td>1.00</td>
<td>0.116</td>
<td>1.356</td>
<td>0.350</td>
<td>0.752</td>
</tr>
<tr>
<td>b2</td>
<td>1.00</td>
<td>0.075</td>
<td>0.589</td>
<td>0.354</td>
<td>0.765</td>
</tr>
</tbody>
</table>

*a* The results exclude three cases in which the objective function did not have a unique minimum.

*b* The results exclude two cases in which the objective function did not have a unique minimum.

### Notes

$Y_{i2} = 0$ = 0.15, $P(Y_{i1} = 0, Y_{i2} > 0) = 0.15$, and $P(Y_{i1} = 0, Y_{i2} = 0) = 0.35$. Except where explicitly noted, these are approximately the censoring probabilities for all the specifications considered here. In the first panel of Table I, it is seen that with this specification, the asymptotic distribution provides a fairly good approximation to the first two moments of the exact distribution of $\hat{\beta}_4$ for sample sizes greater than 200. For a sample size of 20, $\hat{\beta}_4$ is biased (though close to median-unbiased) and its variance and median absolute error are much larger than that suggested by its asymptotic distribution.

The discrepancy between the exact distribution of $\hat{\beta}_4$ and the distribution implied by Theorem 2 is due largely to the finite distribution of $\hat{\beta}_4$ being
asymmetric. Consider, for example, the first coefficient of \( \hat{\beta}_4 \). The 10th, 25th, 50th, 75th, and 90th percentiles of \( \hat{\beta}_4 \) (with those suggested by the asymptotic distribution in parentheses) are: 0.52 (0.53), 0.75 (0.75), 1.01 (1.00), 1.29 (1.25), and 1.64 (1.47). These suggest that with Specification 1, the asymptotic distribution of \( \hat{\beta}_4 \) provides a good approximation for the left half of the exact distribution of (the first coefficient of) \( \hat{\beta}_4 \), whereas the right half of the exact distribution is much thicker than the asymptotic distribution.

The asymmetry of the finite sample distribution of \( \hat{\beta}_4 \) is to be expected. Chesher and Spady (1990) have shown by Monte Carlo experimentation and asymptotic expansions, that the exact distribution of Powell’s trimmed least squares estimator is skewed in small samples. As \( \hat{\beta}_4 \) can be thought of as a bivariate generalization of Powell’s estimator, it is not surprising that the finite sample characteristics of the two estimators are similar.

There are (at least) two reasons to be skeptical about the results from Specification 1. First, it is possible that the normality of the error terms will make \( \hat{\beta}_4 \) look “too good.” If no observations are censored, then \( \hat{\beta}_4 \) reduces to the OLS estimator on the first differences of the observations. As OLS is equivalent to maximum likelihood estimation under normality, this may lead to a performance of \( \hat{\beta}_4 \) that is better than that under other different distributional assumptions. Secondly, joint normality of regressors and error terms in some cases gives consistency and asymptotic normality, where this would otherwise not be the case. Specifications 2 and 3, therefore, make different distributional assumptions.

Specification 2 differs from Specification 1 in the distribution of the errors as well as the regressors. The random variables \( \alpha_i, \eta_{i1}, \eta_{i2}, X_{2i1}, X_{2i2}, e_{i1}, \) and \( e_{i2} \) are independent and all distributed gamma with three degrees of freedom (standardized to have mean 0 and variance 1). The true value of \( \beta \) is (1,1). The censoring probabilities with this specification are \( P(Y_{i1} > 0, Y_{i2} > 0) = 0.30, P(Y_{i1} > 0, Y_{i2} = 0) = 0.16, P(Y_{i1} = 0, Y_{i2} > 0) = 0.16, \) and \( P(Y_{i1} = 0, Y_{i2} = 0) = 0.39. \) In Specification 3, \( \alpha_i, \eta_{i1}, \eta_{i2}, X_{2i1}, X_{2i2}, e_{i1}, \) and \( e_{i2} \) are independent. \( e_{i1} \) and \( e_{i2} \) are normally distributed with mean 0 and variance 1. The remainder are distributed gamma with three degrees of freedom (standardized to have mean 0 and variance 1). Again, the true value of \( \beta \) is (1,1). The censoring probabilities under Specification 3 equal those under Specification 2.

With the exception that the small sample biases seem to be more severe under Specifications 2 and 3, the results do not differ substantially from the results from Specification 1.

Most of the results in Specifications 1 to 3 are for samples that are “unrealistically” large compared to the number of parameters. In Specification 4, we modify Specification 1 by adding three more regressors. The three regressors all have true coefficients of 0, and hence do not change the “R2" of the model. The random variables \( \alpha_i, \eta_{i1}, \eta_{i2}, X_{k1i}, X_{k2i} \) for \( k = 2, \ldots, 5, e_{i1}, \) and \( e_{i2} \) are all independent and normally distributed with mean 0 and variance 1. The true value of \( \beta \) is (1,1,0,0,0). The results of Specification 4 (Table II) are similar to those of Specifications 1, 2, and 3.
### TABLE II
RESULTS FOR DIFFERENT SPECIFICATIONS

<table>
<thead>
<tr>
<th>Specifications</th>
<th>True</th>
<th>Bias</th>
<th>RMSE</th>
<th>ARMSE</th>
<th>LQ</th>
<th>Median</th>
<th>QU</th>
<th>MAE</th>
<th>AMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b₁</td>
<td>1.000</td>
<td>0.005</td>
<td>0.121</td>
<td>0.115</td>
<td>0.925</td>
<td>0.997</td>
<td>1.083</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>b₂</td>
<td>1.000</td>
<td>0.001</td>
<td>0.122</td>
<td>0.115</td>
<td>0.916</td>
<td>1.001</td>
<td>1.078</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>b₃</td>
<td>0.000</td>
<td>-0.004</td>
<td>0.111</td>
<td>0.106</td>
<td>-0.078</td>
<td>-0.004</td>
<td>0.070</td>
<td>0.073</td>
</tr>
<tr>
<td></td>
<td>b₄</td>
<td>0.000</td>
<td>0.003</td>
<td>0.108</td>
<td>0.107</td>
<td>-0.063</td>
<td>0.004</td>
<td>0.071</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>b₅</td>
<td>0.000</td>
<td>-0.003</td>
<td>0.114</td>
<td>0.106</td>
<td>-0.082</td>
<td>-0.003</td>
<td>0.070</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>b₁</td>
<td>1.000</td>
<td>0.009</td>
<td>0.134</td>
<td>0.135</td>
<td>0.919</td>
<td>1.002</td>
<td>1.099</td>
<td>0.087</td>
</tr>
<tr>
<td></td>
<td>b₂</td>
<td>1.000</td>
<td>0.008</td>
<td>0.139</td>
<td>0.135</td>
<td>0.916</td>
<td>1.005</td>
<td>1.093</td>
<td>0.090</td>
</tr>
<tr>
<td></td>
<td>b₃</td>
<td>0.000</td>
<td>-0.008</td>
<td>0.123</td>
<td>0.117</td>
<td>-0.086</td>
<td>-0.003</td>
<td>0.078</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>b₄</td>
<td>0.000</td>
<td>-0.004</td>
<td>0.122</td>
<td>0.117</td>
<td>-0.083</td>
<td>-0.008</td>
<td>0.080</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>b₅</td>
<td>0.000</td>
<td>-0.005</td>
<td>0.120</td>
<td>0.116</td>
<td>-0.084</td>
<td>-0.004</td>
<td>0.074</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>b₁</td>
<td>1.000</td>
<td>0.011</td>
<td>0.122</td>
<td>0.114</td>
<td>0.929</td>
<td>1.008</td>
<td>1.090</td>
<td>0.081</td>
</tr>
<tr>
<td></td>
<td>b₂</td>
<td>1.000</td>
<td>0.007</td>
<td>0.134</td>
<td>0.133</td>
<td>0.914</td>
<td>1.007</td>
<td>1.092</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>b₃</td>
<td>0.000</td>
<td>-0.004</td>
<td>0.105</td>
<td>0.104</td>
<td>-0.074</td>
<td>-0.003</td>
<td>0.069</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>b₄</td>
<td>0.000</td>
<td>0.001</td>
<td>0.105</td>
<td>0.105</td>
<td>-0.062</td>
<td>-0.003</td>
<td>0.064</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>b₅</td>
<td>0.000</td>
<td>0.001</td>
<td>0.108</td>
<td>0.105</td>
<td>-0.065</td>
<td>0.003</td>
<td>0.073</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>b₁</td>
<td>1.000</td>
<td>0.008</td>
<td>0.120</td>
<td>0.117</td>
<td>0.930</td>
<td>1.003</td>
<td>1.084</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>b₂</td>
<td>1.000</td>
<td>0.001</td>
<td>0.161</td>
<td>0.160</td>
<td>0.893</td>
<td>0.992</td>
<td>1.105</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>b₃</td>
<td>0.000</td>
<td>0.008</td>
<td>0.107</td>
<td>0.108</td>
<td>-0.063</td>
<td>0.002</td>
<td>0.079</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>b₄</td>
<td>0.000</td>
<td>0.003</td>
<td>0.109</td>
<td>0.107</td>
<td>-0.071</td>
<td>0.004</td>
<td>0.072</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>b₅</td>
<td>0.000</td>
<td>-0.002</td>
<td>0.111</td>
<td>0.107</td>
<td>-0.081</td>
<td>-0.001</td>
<td>0.073</td>
<td>0.075</td>
</tr>
</tbody>
</table>

In many Monte Carlo studies of nonlinear models, it is found that heteroskedasticity can severely contaminate the estimators. In Specifications 5 and 6, the design in Specification 4 is modified by letting the variance of the error depend on the fixed effect and on one of the regressors, respectively. Specification 5 differs from Specification 4 in that the variances of \( \varepsilon_1 \) and \( \varepsilon_2 \) are \((1/2) + (\alpha^2_i)/2\). In Specification 6 they are \((1/2) + (X_{2i1}^2)/2\). This makes the unconditional variances of the error terms 1, as in Specification 4.

The results presented in Table II indicate that \( \hat{\beta}_4 \) behaves well under heteroskedasticity, and that Theorem 2 provides a good approximation to the small sample distribution of \( \hat{\beta}_4 \) (when the sample size is 200).

Comparing the results of Specifications 1 and 4, the small sample properties of the estimators of the first two elements of \( \beta \) do not seem to depend on the presence of the additional regressors. This may be because the additional regressors are independent of the first two regressors. In Specification 7, we repeat the experiment in Specification 4, but with the additional regressors not being independent of the first two regressors. The difference between Specification 7 and Specification 4 is that in Specification 7, \( X_{3ij} = \nu_{1ij} + \alpha_i \) (so the third
TABLE III

$T$ Statistics Based on $\hat{\beta}_4$; $n = 200$

<table>
<thead>
<tr>
<th>Specification 4.</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (0.20)</td>
<td>0.238</td>
<td>0.223</td>
<td>0.221</td>
<td>0.222</td>
<td>0.246</td>
</tr>
<tr>
<td>Size (0.10)</td>
<td>0.125</td>
<td>0.134</td>
<td>0.131</td>
<td>0.115</td>
<td>0.142</td>
</tr>
<tr>
<td>Size (0.05)</td>
<td>0.072</td>
<td>0.078</td>
<td>0.068</td>
<td>0.060</td>
<td>0.084</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification 5.</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (0.20)</td>
<td>0.209</td>
<td>0.226</td>
<td>0.240</td>
<td>0.211</td>
<td>0.227</td>
</tr>
<tr>
<td>Size (0.10)</td>
<td>0.119</td>
<td>0.118</td>
<td>0.137</td>
<td>0.127</td>
<td>0.112</td>
</tr>
<tr>
<td>Size (0.05)</td>
<td>0.070</td>
<td>0.072</td>
<td>0.077</td>
<td>0.068</td>
<td>0.059</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification 6.</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (0.20)</td>
<td>0.216</td>
<td>0.204</td>
<td>0.223</td>
<td>0.211</td>
<td>0.214</td>
</tr>
<tr>
<td>Size (0.10)</td>
<td>0.114</td>
<td>0.110</td>
<td>0.114</td>
<td>0.122</td>
<td>0.117</td>
</tr>
<tr>
<td>Size (0.05)</td>
<td>0.072</td>
<td>0.073</td>
<td>0.068</td>
<td>0.065</td>
<td>0.064</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification 7.</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
<th>$b_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (0.20)</td>
<td>0.224</td>
<td>0.207</td>
<td>0.219</td>
<td>0.218</td>
<td>0.217</td>
</tr>
<tr>
<td>Size (0.10)</td>
<td>0.133</td>
<td>0.117</td>
<td>0.121</td>
<td>0.124</td>
<td>0.112</td>
</tr>
<tr>
<td>Size (0.05)</td>
<td>0.066</td>
<td>0.063</td>
<td>0.058</td>
<td>0.060</td>
<td>0.056</td>
</tr>
</tbody>
</table>

regressor depends on the fixed effect, and hence is correlated with the first regressor) and $X_{4ij} = v_{2ij} + X_{2ij}$ (so the fourth regressor depends on the second regressor but not on the fixed effect). The random variables $v_{kij}$ are independent with standard normal distributions. The censoring probabilities equal those in Specification 4.

The results for Specification 7 are given in the fourth panel of Table II. There is almost no effect on the first and third coefficients. This is not surprising, as the dependence between the regressors is “differenced out” (though it still might have an effect through the dependent variables). As expected, the estimates of the second coefficient are less precise than in Specification 4. We have no good intuition as to why the precision of the estimates of the fourth coefficient seem to be unchanged from Specification 4 to Specification 7.

Theorems 2 and 3 are of interest mostly because they imply that the usual $t$ statistics are asymptotically normally distributed. For each of the samples used to construct Table II, the estimator of the variance of $\hat{\beta}_4$ was calculated. Based on this, a $t$ statistic that tests whether each coefficient of $\beta$ equals its true value, was calculated. The three lines of each panel of Table III give the fraction of samples for which the true value was rejected at a 20%, 10%, and 5% asymptotic level. These fractions are in the 21–25% range when the asymptotic level is 20%, 11–14% when the asymptotic level is 10%, and between 6 and 8% when the tests are performed at the 5% level. The size of the test exceeds the asymptotic level in all cases. This is due partly to the estimation error in the variance of $\hat{\beta}_4$ and partly to the small sample distribution of $\hat{\beta}_4$ having tails that
are heavier than the asymptotic distribution. For example, when the true (unknown) asymptotic variance of $\hat{\beta}_4$ is used to calculate the $t$ statistic for $\beta_k = 1$, the rejection frequency falls in all of the cases considered in Table III, but it remains above the asymptotic level in 53 out of the 60 cases.

7. CONCLUDING REMARKS

This paper has suggested estimators of truncated and censored regression models with fixed effects. The idea behind the estimators is to exploit a symmetry in the distribution of the latent variables. If the true value of the parameter was known, then trimming could introduce the same symmetry in the distribution of the observed variables. This suggests orthogonality conditions which must hold at the true value of the parameter. These orthogonality conditions are the asymptotic first order conditions for minimization of the objective functions used to define the estimators in this paper.

The estimators are proved to be consistent and asymptotically normal, and consistent estimators of the asymptotic variances are presented. A small Monte Carlo experiment investigates the small sample distribution of one of the estimators. The results give some indication that the asymptotic distribution is a good approximation of the small sample distribution when the sample size is 200 or larger, but that the small sample distribution of the estimator is skewed when the sample size is small.

The estimators presented in this paper all assume that the data consist of a panel of $n$ individuals, each observed at two points in time. However, most panel data sets have more than two observations per individual. The estimators presented here can easily be modified to cover this case as well. Suppose that we want to use $\hat{\beta}_4$ in a case where there are $T \geq 2$ observations for each individual (for simplicity assume that $T$ is nonrandom and constant across individuals). If $\{e_{it}\}_{t=1}^T$ are i.i.d., then equation (2.5) must hold for all pairs of observations $s, t$ (because $T_n$ is convex, $\beta$ will be the unique vector that satisfies the orthogonality conditions). This gives $KT(T-1)/2$ orthogonality conditions that can be used to estimate $\beta$. Theorem 3.2 of Hansen (1982) can be used to construct the optimal weights to give to the orthogonality conditions. The optimal weights will depend on the true value of $\beta$, and an initial estimator will, therefore, be needed. One possible choice of initial estimator is one that gives equal weight to all the pairs of observations:

$$\hat{\beta}^*_4 = \arg\min_{b \in B} \sum_{i=1}^n \sum_{s < t} \chi(Y_{is}, Y_{it}, (X_{is} - X_{it})b)$$

It is clear that the symmetry exploited in Section 2 can be used to derive an infinite number of orthogonality conditions. For example, for the censored regression model, (2.2)–(2.5) must all hold. Theorem 3.2 of Hansen (1982) can be used to optimally combine a finite number of these orthogonality conditions. In future research, we plan to explore estimation strategies that use all information contained in Proposition 1. It is also clear that it is possible to derive
estimators based on loss functions other than the absolute or quadratic loss. A library of GAUSS-routines that calculates the estimators for the censored regression model with fixed effects from a panel data set of arbitrary length is available from the author.

Another topic with which we have not dealt here, is the question of lagged endogenous variables. In the Monte Carlo study cited in the Introduction, Heckman (1979) found that the fixed effects probit estimator performs much worse when there are lagged endogenous variables. It would therefore be interesting to investigate whether "trimmed" estimators like those in this paper can be constructed for models with lagged endogenous variables. This issue is pursued in Honoré (1990).

Department of Economics, Northwestern University, Evanston, IL 60208, U.S.A.

APPENDIX 1: LAD INTERPRETATION OF $\hat{\beta}_1$

In this appendix, we show that the estimator $\hat{\beta}_1$ can be interpreted as a least absolute deviations estimator. Suppose that the data are from the truncated model, and that we are to estimate $\beta$ and $\{a_i\}_{i=1}^n$ by Powell's least absolute deviations estimator for censored regression:

$$
\min_{b(A_i; i=1, \ldots, n)} \sum_{i=1}^n \sum_{t=1}^2 |Y_{it} - \max \{0, a_i + X_{it} b\}|.
$$

Below, it is proved that the solution for $b$ would be given by minimization of $Q_n$. Note that this is the minimization used to find the least absolute deviations estimator for the censored regression model—not for the truncated regression model.

It will be proved that

$$
\min_a \sum_{t=1}^2 |Y_{it} - \max \{0, a + X_{it} b\}| = \psi(Y_{i1}, Y_{i2}, (X_{i1} - X_{i2}) b),
$$

which establishes the desired result.

Let $b$ be given, and consider the $i$th individual. Without loss of generality, assume that $Y_{i1} - X_{i1} b > Y_{i2} - X_{i2} b$. There are two cases to be considered.

First, consider the case where $Y_{i2} - X_{i2} b > -X_{i1} b$. The function

$$
g(a, b) = \sum_{t=1}^2 |Y_{it} - \max \{0, a + X_{it} b\}|
$$

is then continuous and piecewise linear with slope

$$
\frac{\partial g}{\partial a} = \begin{cases} 
0 & \text{for } a < \min \{-X_{i1} b, -X_{i2} b\}, \\
-1 & \text{for } \min \{-X_{i1} b, -X_{i2} b\} < a < \max \{-X_{i1} b, -X_{i2} b\} \\
-2 & \text{for } \max \{-X_{i1} b, -X_{i2} b\} < a < Y_{i2} - X_{i2} b, \\
0 & \text{for } Y_{i2} - X_{i2} b < a < Y_{i1} - X_{i2} b, \\
2 & \text{for } a > Y_{i1} - X_{i1} b.
\end{cases}
$$

The minimum is therefore attained in the interval $Y_{i2} - X_{i2} b < a < Y_{i1} - X_{i1} b$, and the minimized value is $(Y_{i1} - X_{i1} b) - (Y_{i2} - X_{i2} b)$, as was to be shown.
Second, consider the case where \( Y_{i2} - X_{i2}b < -X_{i1}b \). The function

\[
g(a, b) = \sum_{i=1}^{2} |Y_{ii} - \max\{0, a + X_{ii}b\}|
\]

is then again continuous and piecewise linear. Its slope

\[
\frac{\partial g}{\partial a} = \begin{cases} 
0 & \text{for } a < -X_{i2}b, \\
-1 & \text{for } -X_{i2}b < a < Y_{i2} - X_{i2}b, \\
1 & \text{for } Y_{i2} - X_{i2}b < a < -X_{i1}b, \\
0 & \text{for } -X_{i1}b < a < Y_{i1} - X_{i1}b, \\
2 & \text{for } a > Y_{i1} - X_{i1}b.
\end{cases}
\]

The minimum of \( \sum_{i=1}^{2} |Y_{ii} - \max\{0, a + X_{ii}b\}| \) is then attained at \( a = Y_{i2} - X_{i2}b \), and the value is then \( Y_{i1} \).

**APPENDIX 2: SOME USEFUL LEMMAS**

This Appendix states and proves a number of lemmas which will be used to prove the theorems stated in the text. Let

\[
\psi^\circ(z_1, z_2, \delta, \delta_0) = 1\{z_1 > 0, z_2 > 0\}(\psi(z_1, z_2, \delta) - \psi(z_1, z_2, \delta_0)),
\]

\[
\xi^\circ(z_1, z_2, \delta, \delta_0) = 1\{z_1 > 0, z_2 > 0\}(\psi(z_1, z_2, \delta)^2 - \psi(z_1, z_2, \delta_0)^2),
\]

\[
\varphi^\circ(z_1, z_2, \delta, \delta_0) = \varphi(\max\{0, z_1\}, \max\{0, z_2\}, \delta) - \varphi(\max\{0, z_1\}, \max\{0, z_2\}, \delta_0)
\]

and

\[
\chi^\circ(z_1, z_2, \delta, \delta_0) = \chi(\max\{0, z_1\}, \max\{0, z_2\}, \delta) - \chi(\max\{0, z_1\}, \max\{0, z_2\}, \delta_0).
\]

Also let

\[
Q_n^\circ(b) = \frac{1}{n}(Q_n(b) - Q_n(\beta)) = \frac{1}{n} \sum_{i=1}^{n} \psi^\circ(Y_{i1}, Y_{i2}, \Delta X_i b, \Delta X_i \beta),
\]

\[
R_n^\circ(b) = \frac{1}{n}(R_n(b) - R_n(\beta)) = \frac{1}{n} \sum_{i=1}^{n} \xi^\circ(Y_{i1}, Y_{i2}, \Delta X_i b, \Delta X_i \beta),
\]

\[
S_n^\circ(b) = \frac{1}{n}(S_n(b) - S_n(\beta)) = \frac{1}{n} \sum_{i=1}^{n} \varphi^\circ(Y_{i1}, Y_{i2}, \Delta X_i b, \Delta X_i \beta),
\]

and

\[
T_n^\circ(b) = \frac{1}{n}(T_n(b) - T_n(\beta)) = \frac{1}{n} \sum_{i=1}^{n} \chi^\circ(Y_{i1}, Y_{i2}, \Delta X_i b, \Delta X_i \beta).
\]

\( \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \) and \( \hat{\beta}_4 \) are then defined as the minimizers of \( Q_n^\circ, R_n^\circ, \) and \( T_n^\circ, \) respectively. There are two reasons for working with \( Q_n^\circ \) through \( T_n^\circ \), rather than \( Q_n \) through \( T_n \). The first is that the assumptions needed for existence of moments of \( Q_n^\circ \) through \( T_n^\circ \) are weaker than those needed for existence of moments of \( Q_n \) through \( T_n \). The second reason is that \( \psi^\circ \) through \( \chi^\circ \) are defined for negative values of \( z_1 \) and \( z_2 \) in such a way that the minimizations of \( Q_n^\circ \) through \( T_n^\circ \) are unchanged if the truncated or censored observations are replaced with their latent counterparts.
Lemma A.1: Let \((Z_1, Z_2)\) be continuously distributed with density \(f\). Then

\[
\frac{\partial E[\psi^0(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} = E[\{Z_1 - Z_2 < \delta < Z_1, Z_1 > 0, Z_2 > 0\}]
- E[\{Z_2 - Z_1 < -\delta < Z_2, Z_1 > 0, Z_2 > 0\}]
\]

and

\[
\frac{\partial E[\phi^0(Z_1, Z_2, \delta)]}{\partial \delta} = E[\{Z_2 > 0, Z_2 > \max\{0, Z_1 - \delta\}\}]
- E[\{Z_1 > 0, Z_1 > \max\{0, Z_2 + \delta\}\}].
\]

If, in addition, \((Z_1, Z_2)\) has finite first moments, then

\[
\frac{\partial E[\xi^\alpha(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} = -2E[\{Z_1 > 0, Z_2 > 0, -Z_2 < \delta < Z_1\}(Z_1 - Z_2 - \delta)]
\]

and

\[
\frac{\partial E[\chi^\xi(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} = 2E[\{\delta > \max\{0, Z_1\}\} \max\{0, Z_2\} - 1\{\delta < -\max\{0, Z_2\}\} \max\{0, Z_1\}]
+ 1\{-\max\{0, Z_2\} < \delta < \max\{0, Z_1\}\}(\max\{0, Z_2\} - \max\{0, Z_1 + \delta\}].
\]

The last statement is true, even if \((Z_1, Z_2)\) is not continuously distributed.

Proof: In all cases, use the Lebesgue dominated convergence theorem to justify differentiation under expectation. Q.E.D.

Lemma A.2: Assume that \((Z_1, Z_2)\) is a continuously distributed random vector, such that \(P(Z_1 > 0, Z_2 > 0) > 0\) and the distribution of \(Z_1 - Z_2 - \delta_0\) conditional on \(Z_1 + Z_2\) is symmetric around \(0\) for some real number \(\delta_0\).

(a) If the density of \(Z_1 - Z_2 - \delta_0\) conditional on \(Z_1 + Z_2\) is strictly unimodal, then \(E[\psi^0(Z_1, Z_2, \delta, \delta_0)]\) is uniquely minimized at \(\delta = \delta_0\).

(b) If the density of \(Z_1 - Z_2 - \delta_0\) conditional on \(Z_1 + Z_2\) is strictly unimodal, and \((Z_1, Z_2)\) has finite first moments, then \(E[\xi^\alpha(Z_1, Z_2, \delta, \delta_0)]\) is uniquely minimized at \(\delta = \delta_0\).

(c) \(E[\phi^0(Z_1, Z_2, \delta, \delta_0)]\) is uniquely minimized at \(\delta = \delta_0\).

(d) If \((Z_1, Z_2)\) has finite first moments, then \(E[\chi^\xi(Z_1, Z_2, \delta, \delta_0)]\) is uniquely minimized at \(\delta = \delta_0\) (this is true even if \((Z_1, Z_2)\) is not continuously distributed).

If \(P(Z_1 > 0, Z_2 > 0) = 0\), then (a)–(d) still hold, except that the \(\delta = \delta_0\) need not be the unique minimizer.

Proof: Let \(\delta_0\) be given as in the statement of the lemma and let \(\delta\) be an arbitrary real number. Define \(W_1\) and \(W_2\), by \(W_1 = Z_1 - Z_2 - \delta_0\) and \(W_2 = Z_1 + Z_2 - |\delta|\). Also define \(\eta = \delta_0 - \delta\).

We can then write

\[
\frac{\partial E[\psi^0(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} = E[\{W_2 > 0, W_2 < W_1 + \eta < 0\}] - E[\{W_2 > 0, 0 < W_1 + \eta < W_2\}] \\
= E[\{W_2 > 0, -W_2 - \eta < W_1 < -\eta\}] - E[\{W_2 > 0, -\eta < W_1 < W_2 - \eta\}] \\
= E[\{W_2 > 0\} P(-W_2 - \eta < W_1 < -\eta | W_2) - P(-\eta < W_1 < W_2 - \eta | W_2)]
\]

As the conditional distribution of \(W_1\) given \(W_2\) is unimodal and symmetric around \(0\), this is negative for \(\delta < \delta_0\) (\(\eta > 0\), 0 at \(\delta = \delta_0\) (\(\eta = 0\)), and positive for \(\delta > \delta_0\) (\(\eta < 0\)). Therefore, \(E[\psi^0(Z_1, Z_2, \delta, \delta_0)]\) is minimized at \(\delta = \delta_0\). This proves (a).
By the same line of reasoning

$$\frac{\partial E[\xi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} = -2E[1\{W_2 < W_1 + \eta < W_2\}(W_1 + \eta)].$$

As $W_1 + \eta$ is unimodal and symmetric around $\eta$, this is negative for $\delta < \delta_0$ ($\eta > 0$), 0 at $\delta = \delta_0$ ($\eta = 0$), and positive for $\delta > \delta_0$ ($\eta < 0$). Therefore $E[\xi^*(Z_1, Z_2, \delta, \delta_0)]$ is minimized at $\delta = \delta_0$.

This proves (b).

To prove (c) and (d), note that the assumptions of the lemma imply that $(Z_1, Z_2 + \delta_0)$ is distributed like $(Z_2 + \delta_0, Z_1)$. This, in turn, implies that

$$\frac{\partial E[\phi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} \bigg|_{\delta = \delta_0} = 0 \quad \text{and} \quad \frac{\partial E[\chi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta} \bigg|_{\delta = \delta_0} = 0.$$

Since both of these derivatives are increasing as a function of $\delta$, it follows that $E[\phi^*(Z_1, Z_2, \delta, \delta_0)]$ and $E[\chi^*(Z_1, Z_2, \delta, \delta_0)]$ are minimized at $\delta = \delta_0$. This completes the proof. Q.E.D.

**LEMMA A.3:** Assume that $(Z_1, Z_2)$ is distributed symmetrically around the 45°-line through $(\delta_0, 0)$ (or, equivalently, through $(0, -\delta_0)$).

If $(Z_1, Z_2)$ is continuously distributed with density $f$ then

$$\frac{\partial^2 E[\phi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \bigg|_{\delta = \delta_0} = \begin{cases} P(Z_1 > 0, Z_2 > 0) & \left(2f_{Z_1,Z_2}(\delta_0 | Z_1 > 0, Z_2 > 0) - f_{Z_2}(\delta_0 | Z_1 > 0, Z_2 > 0)\right) \\ P(Z_1 > 0, Z_2 > 0) & \left(2f_{Z_1,Z_2}(\delta_0 | Z_1 > 0, Z_2 > 0) - f_{Z_2}(-\delta_0 | Z_1 > 0, Z_2 > 0)\right) \end{cases}$$

and

$$\frac{\partial^2 E[\phi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \bigg|_{\delta = \delta_0} = \begin{cases} 2 \int_0^\infty f(\delta_0 + z_2, z_2) \, dz_2 + \int_{-\infty}^0 f(\delta_0, z_2) \, dz_2, & \text{for } \delta_0 > 0, \\ 2 \int_0^\infty f(z_1, z_1 - \delta_0) \, dz_1 + \int_{-\infty}^0 f(z_1, -\delta_0) \, dz_1, & \text{for } \delta_0 < 0. \end{cases}$$

If $(Z_1, Z_2)$ is continuously distributed with density $f$ and finite first moments, then

$$\frac{\partial^2 E[\chi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \bigg|_{\delta = \delta_0} = \begin{cases} 2P(Z_1 > \delta_0, Z_2 > 0) - 2\int_0^\infty z_2 f(\delta_0, z_2) \, dz_2 & \text{for } \delta_0 > 0, \\ 2P(Z_2 > -\delta_0, Z_1 > 0) - 2\int_0^\infty z_1 f(z_1, -\delta_0) \, dz_1 & \text{for } \delta_0 < 0, \end{cases}$$

and

$$\frac{\partial^2 E[\chi^*(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \bigg|_{\delta = \delta_0} = 2P(-\max\{0, Z_2\} < \delta_0 < \max\{0, Z_1\}).$$

**PROOF:** It is simple to find the second derivative of $E[\chi^*(Z_1, Z_2, \delta, \delta_0)]$ and the second derivatives of $E[\phi^*(Z_1, Z_2, \delta, \delta_0)]$, $E[\phi^*(Z_1, Z_2, \delta, \delta_0)]$, and $E[\chi^*(Z_1, Z_2, \delta, \delta_0)]$ evaluated at any
point except 0. The difficulty in the proof arises because the second derivatives might not exist at \( \delta = 0 \). However, the first derivatives (given in Lemma A.1) have right and left derivatives given by

\[
\left[ \frac{\partial^2 E[\psi^0(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \right]_+
\]

\[
= \begin{cases} 
  P(Z_1 > 0, Z_2 > 0)(2f_{Z_1 - Z_2}(\delta | Z_1 > 0, Z_2 > 0) - f_{Z_1}(\delta | Z_1 > 0, Z_2 > 0)) 
  & \text{for } \delta > 0, \\
  P(Z_1 > 0, Z_2 > 0)(2f_{Z_1 - Z_2}(\delta | Z_1 > 0, Z_2 > 0) - f_{Z_2}(\delta | Z_1 > 0, Z_2 > 0)) 
  & \text{for } \delta < 0, 
\end{cases}
\]

\[
\left[ \frac{\partial^2 E[\psi^0(Z_1, Z_2, \delta, \delta_0)]}{\partial \delta^2} \right]_-
\]

\[
= \begin{cases} 
  P(Z_1 > 0, Z_2 > 0)(2f_{Z_1 - Z_2}(\delta | Z_1 > 0, Z_2 > 0) - f_{Z_2}(\delta | Z_1 > 0, Z_2 > 0)) 
  & \text{for } \delta > 0, \\
  P(Z_1 > 0, Z_2 > 0)(2f_{Z_1 - Z_2}(\delta | Z_1 > 0, Z_2 > 0) - f_{Z_1}(\delta | Z_1 > 0, Z_2 > 0)) 
  & \text{for } \delta < 0, 
\end{cases}
\]

The left and right derivatives will, in general, be different at \( \delta = 0 \). However, if \( \delta_0 = 0 \), \( f(z_1, z_2) = f(z_2, z_1) \) and the left and right derivatives agree at \( \delta = 0 \).

**Q.E.D.**

**LEMMA B.1 (Uniform Convergence):** Let \( B \) be a compact set with the true parameter value, \( \beta \), in the interior (for \( B_1 \) and \( B_2 \), \( B \) can be thought of as the parameter space). Under the assumptions of Theorem 1, \( Q^0_n \), \( R^0_n \), \( S^0_n \), and \( T^0_n \) converge to their expectations uniformly in \( b \in B \).

I thank an anonymous referee for pointing out a flaw in the use of an earlier version of this lemma, and for suggesting the current version.
PROOF: First, notice that \( \psi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta), \chi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta) \) and \( \delta(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta) \) are measurable functions of \( (\alpha_1, \alpha_2, X_1, X_2) \) for a given \( b \), and continuous functions of \( b \) given \( (\alpha_1, \alpha_2, X_1, X_2) \).

From the definitions of \( \psi, \varphi, \) and \( \chi \) (see also Figures 2–5, which makes the following expressions trivial), it follows that

\[
\begin{align*}
|\psi(z_1, z_2, \delta_1) - \psi(z_1, z_2, \delta_2)| & \leq |\delta_1 - \delta_2|, \\
|\varphi(z_1, z_2, \delta_1)^2 - \varphi(z_1, z_2, \delta_2)^2| & \leq 2 \max\{z_1, z_2\}|\delta_1 - \delta_2|, \\
|\chi(z_1, z_2, \delta_1) - \chi(z_1, z_2, \delta_2)| & \leq |\delta_1 - \delta_2|, \\
|\chi(z_1, z_2, \delta_1) - \chi(z_1, z_2, \delta_2)| & \leq 2 \max\{z_1, z_2\}|\delta_1 - \delta_2|. 
\end{align*}
\]

This implies that

\[
E\left[ \sup_{b \in B} |\psi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta)| \right] < \infty, \\
E\left[ \sup_{b \in B} |\varphi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta)| \right] < \infty, \\
E\left[ \sup_{b \in B} |\chi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta)| \right] < \infty,
\]

and

\[
E\left[ \sup_{b \in B} |\chi(\alpha_1, \alpha_2, \Delta X_1 b, \Delta X_1 \beta)| \right] < \infty,
\]

where Assumption M.1 is used in the first and third equations, and Assumption M.2 is used in the second and fourth equations.

The result now follows from a standard uniform law of large numbers (for example, Amemiya (1985), Theorem 4.2.1). Q.E.D.

**LEMMA B.2 (Identification):** Under the assumptions of Theorem 1, \( E[Q_n^o(b)], E[R_n^o(b)], E[S_n^o(b)], \) and \( E[T_n^o(b)] \) are uniquely minimized at \( b = \beta \).

**PROOF:** First notice that

\[
E[\psi(\alpha, \Delta X b, \Delta X \beta)] = E\left[ E[\psi(\alpha, \Delta X b, \Delta X \beta) | \alpha, X_1, X_2] \right].
\]

Consider \( \alpha, x_1, x_2 \) in the support of \( \alpha, X_1, X_2 \) and let \( \delta_0 = (x_1 - x_2)\beta \). Lemma A.2 then implies that \( E[\psi(\alpha, \Delta X b, \Delta X \beta)] \) is minimized at a \( b \) such that \( \Delta X b = \Delta X \beta \) with probability 1. Assumption R.1 implies that the only \( b \) satisfying this is \( b = \beta \). This proves the first statement of the lemma. The remaining statements can be proved in exactly the same way. Q.E.D.

**APPENDIX 3: PROOFS OF THEOREMS AND LEMMAS IN THE TEXT**

**PROOF OF THEOREM 1:** Rather than proving that \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are consistent as the number of observations for which \( Y_1^* > 0 \) and \( Y_2^* > 0 \) approaches infinity, we will prove that they are consistent as the number of latent observations increases (the latter clearly implies the former). In this proof, \( n \) will therefore denote the number of observations in the latent sample. This makes the notation a little clearer, because it will not be necessary to work with distributions conditional on \( Y_1^* > 0 \) and \( Y_2^* > 0 \). As \( \psi(z_1, z_2, \delta, \delta_0) = 0 \), whenever either \( z_1 < 0 \) or \( z_2 < 0 \), the estimators will not change, if the observations for which either \( Y_1^* \leq 0 \) or \( Y_2^* \leq 0 \) are included in the definition of \( Q_n^o \) and \( R_n^o \).

To prove consistency of \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \), we will verify the conditions of Theorem 4.1.1 in Amemiya (1985). Assumption A (compactness) is satisfied by Assumption P.1; Assumption B (continuity and measurability) is trivially satisfied; finally, Assumption C (uniform convergence and identification) is satisfied by Lemmas B.1 and B.2.
For $\hat{\beta}_3$ and $\hat{\beta}_4$, it suffices to verify the conditions of Lemma A in Newey and Powell (1987). Assumption A (uniform convergence) is satisfied by Lemma B.1, Assumption B (identification) is satisfied by Lemma B.2, and Assumption C (convexity) is trivially verified. This proves that $\hat{\beta}_3$ and $\hat{\beta}_4$ are consistent. Q.E.D.

**Proof of Lemma 1:** The symmetry part is obvious. The unimodality is proved as follows: Let $f$ be the (common) marginal density for $\varepsilon_1$ and $\varepsilon_2$. The unimodality in the statement of the Lemma is then implied by $f(\varepsilon - u)f(\varepsilon + u)$ being strictly increasing for $u < 0$ (and hence decreasing for $u > 0$) for all $\varepsilon$, or

$$f(\varepsilon - u_1)f(\varepsilon + u_1) > f(\varepsilon - u_2)f(\varepsilon + u_2)$$

or

$$\log f(\varepsilon - u_1) + \log f(\varepsilon + u_1) > \log f(\varepsilon - u_2) + \log f(\varepsilon + u_2)$$

for all $\varepsilon$ and $|u_2| < |u_1|$ such that $\varepsilon - u_1, \varepsilon + u_1, \varepsilon - u_2$, and $\varepsilon + u_2$ are in the support of $f$. It is easily seen that this is true if $\log f$ is strictly concave. Q.E.D.

**Proof of Theorem 2:** For notational convenience let $Z = (\varepsilon_1, \varepsilon_2, X_1, X_2, \alpha)$. Define $V_1^* = P(Y_1 > 0, Y_2 > 0)\mathbb{1}, V_2^* = P(Y_1 > 0, Y_2 > 0)\mathbb{1}, \Gamma_1^* = P(Y_1 > 0, Y_2 > 0)\mathbb{1}$, and $\Gamma_2^* = P(Y_1 > 0, Y_2 > 0)\mathbb{1}$. The statements in (i) and (ii) are then implied by

$$\sqrt{n} (\hat{\beta}_1 - \beta) \rightarrow N\left(0, \Gamma_1^{-1} V_1^* \Gamma_1^{-1}\right)$$

and

$$\sqrt{n} (\hat{\beta}_2 - \beta) \rightarrow N\left(0, \Gamma_2^{-1} V_2^* \Gamma_2^{-1}\right),$$

where $n$ now denotes the number of observations in the latent sample.

First, note that the assumptions of Theorem 1 are satisfied, so $\hat{\beta}_l$, $l = 1, \ldots, 4$, are consistent. Asymptotic normality will be proved by defining functions, $f_l(Z, b)$, and then verifying that for $l = 1, \ldots, 4, G_l$ and $G_{ln}$ defined by

$$G_l(b) = E[f_l(Z, b)] \quad \text{and} \quad G_{ln}(b) = \frac{1}{n} \sum_{i=1}^{n} f_l(Z_i, b)$$

satisfy:

(i) $\|G_{ln}(\hat{\beta}_l)\| = o_p(n^{-1/2}).$

(ii) $G_l(\cdot)$ is differentiable at $\beta$ with derivative matrix $\Gamma_l$ ($\Gamma_l^*$ for $l = 1, 2$) of full rank.

(iii) For any sequence $(\delta_n)$ of positive numbers such that $\delta_n \to 0$ as $n \to \infty$,

$$\sup_{\|b - \beta\| < \delta_n} \|G_{ln}(b) - G_l(b) - G_{ln}(\beta)\| = o_p(n^{-1/2}).$$

(iv) $\sqrt{n} G_{ln}(\beta) \rightarrow N(0, V_l^*)$ ($N(0, V_l^*)$ for $l = 1, 2$).

(v) $\beta$ is an interior point of the parameter space.

We can then use Theorem 3.3 of Pakes and Pollard (1989) (and the consistency of $\hat{\beta}_l$) to conclude that

$$\sqrt{n} (\hat{\beta}_l - \beta) \rightarrow N\left(0, \Gamma_l^{-1} V_l \Gamma_l^{-1}\right).$$

Under our assumptions, $\Gamma_l$ is symmetric and of full rank, so this expression reduces to

$$\sqrt{n} (\hat{\beta}_l - \beta) \rightarrow N\left(0, \Gamma_l^{-1} \Gamma_l^{-1}\right).$$

Define

$$f_1(Z, b) = E[1\{Y_1^* > 0, Y_2^* > 0\} \{1(\Delta Y^* < \Delta X b < Y_1^*) - 1(-Y_2^* < \Delta X b < \Delta Y^*)\} \Delta X],$$

$$f_2(Z, b) = -E[1\{Y_1^* > 0, Y_2^* > 0\} \{1(-Y_2^* < \Delta X b < Y_1^*) (Y_2^* - Y_2^* - \Delta X b) \Delta X],$$

$$f_3(Z, b) = E[1\{Y_1^* > 0, Y_2^* > 0\} \{1(\Delta X b > \max\{0, Y_1^*\}) \max\{0, Y_2^*\} - \Delta X b\} - 1(-\max\{0, Y_2^*\} > \max\{0, Y_2^*\} + \Delta X b) \Delta X],$$

$$f_4(b, Z) = E[1\{\Delta X b > \max\{0, Y_1^*\}\} \max\{0, Y_2^*\} - 1(\Delta X b > \max\{0, Y_2^*\}) \max\{0, Y_1^*\}] + E[1(-\max\{0, Y_2^*\} < \Delta X b < \max\{0, Y_1^*\}) \times (\max\{0, Y_2^*\} - \max\{0, Y_1^*\} + \Delta X b) \Delta X].$$

BO E. HONORÉ
With these definitions, \( f_1, f_2, f_3, \) and \( f_4 \) are the derivatives of \( \psi^c(Y_1^*, Y_2^*, \Delta Xb, \Delta X\beta) \), \( \phi^c(Y_1^*, Y_2^*, \Delta Xb, \Delta X\beta)/2 \), \( \phi^c(Y_1^*, Y_2^*, \Delta Xb, \Delta X\beta) \), and \( \chi^c(Y_1^*, Y_2^*, \Delta Xb, \Delta X\beta)/2 \) respectively, at all points at which these are differentiable. Therefore \( G_{in}^c (i = 1, \ldots, 4) \) are the derivatives (at all points of differentiability) of \( Q_n(b), R_n(b), T_n(b), \) and \( S_n(b) \).

As noted earlier, \( S_n \) is everywhere differentiable so \( G_{4n}(\hat{b}_4) = 0 \). This proves (i) for \( \hat{b}_4 \). To prove (i) for \( \hat{b}_1 \), observe that the left and right derivatives of \( Q_n(b) \) with respect to the \( k \)th coordinate of \( b, Q_{k,n}(\hat{b}_1) \) and \( Q_{k,n}(\hat{b}_1) \), must satisfy \( Q_{k,n}(\hat{b}_1) \leq 0 \) and \( Q_{k,n}(\hat{b}_1) \geq 0 \). But \( Q_{k,n}(\hat{b}_1) \) and \( Q_{k,n}(\hat{b}_1) \) equal the \( k \)th coordinate of \( G_{in}(\hat{b}_1), G_{k,n}(\hat{b}_1), \) plus the contribution to the left and right derivatives from observations that are located at the kinks of \( \psi(Y_1^*, Y_2^*, \Delta X\beta) \). Hence, the absolute value of the \( k \)th coordinate of \( G_{4n}(\hat{b}_1) \) cannot exceed the absolute value of the maximum contribution of the observations at the kinks to the left and right derivatives of \( Q_n(b) \). Therefore

\[
|G_{k,1n}(\hat{b}_1)| \leq \max_{1 \leq i \leq n} \left\{ \left| \Delta X_i \right| \right\} \times \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta X_i \hat{b}_1 = -Y_i^*, \Delta X_i \neq 0 \right\} + \left\{ \Delta X_i \hat{b}_1 = Y_i^*, \Delta X_i \neq 0 \right\}.
\]

By the same argument

\[
|G_{2,2n}(\hat{b}_2)| \leq \max_{1 \leq i \leq n} \left\{ \left| \Delta X_i \right| \right\} \times \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta X_i \hat{b}_2 = -Y_i^*, \Delta X_i \neq 0 \right\} + \max_{1 \leq i \leq n} \left\{ \left| \Delta X_i \right| \right\} \times \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta X_i \hat{b}_2 = Y_i^*, \Delta X_i \neq 0 \right\}
\]

and

\[
|G_{k,3n}(\hat{b}_3)| \leq \max_{1 \leq i \leq n} \left\{ \left| \Delta X_i \right| \right\} \times \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta X_i > 0, \Delta X_i \hat{b}_3 = Y_i^* - \max \left\{ 0, Y_i^* \right\} \right\} + \left\{ \Delta X_i > 0, -\Delta X_i \hat{b}_3 = Y_i^* - \max \left\{ 0, Y_i^* \right\} \right\}.
\]

To see that (i) is satisfied, observe that (a) the sums on the right-hand side of these inequalities are all \( O_p(1) \) (because \( (Y_1^*, Y_2^*) \) is continuously distributed), and (b) \( \max_{1 \leq i \leq n} \left\{ \left| \Delta X_i \right| \right\} = o_p(\sqrt{n}) \) (because \( E[\left| \Delta X \right|^2] < \infty \)) and \( \max_{1 \leq i \leq n} \left\{ Y_i^* \left| \Delta X_i \right| \right\} = o_p(\sqrt{n}) \) (because \( E[Y_i^* \left| \Delta X \right|^2] < \infty \)). This completes the proof on (i).\(^{10}\)

To verify (ii) for \( \hat{b}_1 \), note that

\[
\begin{align*}
\hat{G}_1(b) &= E \left[ (P(AY^* < AXb < Y_1^*, Y_2^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \
&\quad - P(-Y_2^* < AXb < -AY^*, Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) ) \right] \Delta X.
\end{align*}
\]

Using Lemma A.3, we have for almost all \((X_1, X_2, \alpha)\)

\[
\left| \frac{\partial}{\partial b_j} (P(AY < AXb < Y_1^*, Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \
&\quad - P(-Y_2^* < AXb < -AY, Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha)) \right| \Delta X_k
\]

\[
< \left| (2f_{Y_1^*-Y_2^*}(AXb | X_1, X_2, \alpha) + f_{Y_1^*}(AXb | X_1, X_2, \alpha) \
&\quad + f_{Y_2^*}(-AXb | X_1, X_2, \alpha) ) \right| \Delta X_j \Delta X_k
\]

\[
= \left| (2f_{Y_1^*-Y_2^*}(AX(b-\beta) | X_1, X_2, \alpha) + f_{Y_1^*}(X_1(b-\beta) - X_2b - \alpha | X_1, X_2, \alpha) \right| \Delta X_j \Delta X_k
\]

\[
< |K \Delta X_j \Delta X_k|
\]

\(^{10}\) The proof of (i) for \( l = 1, 2, 3 \) is similar to the proofs in Ruppert and Carroll (1980) and in Powell (1984, 1986).
for some constant $K$ (by Assumption E.4). This means that we can find the derivative of $G$ by differentiating under the expectation on the right-hand side of (A1) (see Cramér (1945, pp. 67–68)):

$$
\frac{\partial G_{k,1}(\beta)}{\partial \beta_j} = E\left[ P(Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \times \left( 2f_{Y_1^*-Y_2^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right) \right. \\
\left. \quad -1(\Delta X \beta \geq 0)f_{Y_2^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right] \\
\left. \quad -1(\Delta X \beta < 0)f_{Y_2^*}(-\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \Delta X_j \Delta X_k \right]
$$

$$
= P(Y_1^* > 0, Y_2^* > 0) \times E\left[ \left( 2f_{Y_1^*-Y_2^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right) \right. \\
\left. \quad -1(\Delta X \beta \geq 0)f_{Y_2^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right] \\
\left. \quad -1(\Delta X \beta < 0)f_{Y_2^*}(-\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right] \Delta X_j \Delta X_k \\
\times \Delta X_j \Delta X_k | Y_1^* > 0, Y_2^* > 0 \right].
$$

Therefore, (ii) is satisfied and $I^*_1$ has full rank by Assumption V.1. This proves (ii) for $l = 1$. For $l = 2$ note that

$$
G_2(b) = -E\left[ P(-Y_2^* < \Delta X b < Y_1^*, Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \times E\left[ Y_1^* - Y_2^* - \Delta X b | -Y_2^* < \Delta X b < Y_1^*, Y_1^* > 0, Y_2^* > 0, X_1, X_2, \alpha \right] \Delta X \right]
$$

and that, using Lemma A.3 and Assumption E.5,

$$
\left| \frac{\partial}{\partial b_j} P(-Y_2^* < \Delta X b < Y_1^*, Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \times E\left[ Y_1^* - Y_2^* - \Delta X b | -Y_2^* < \Delta X b < Y_1^*, Y_1^* > 0, Y_2^* > 0, X_1, X_2, \alpha \right] \Delta X \right|
$$

$$
\leq \left( 1 + 1(\Delta X b \geq 0) \right) \int_{y_1}^{\infty} f_{Y_2^*, Y_2^*}(\Delta X b, y_2 | X_1, X_2, \alpha) dy_2 \\
+ 1(\Delta X b < 0) \int_{0}^{\infty} f_{Y_1^*, Y_2^*}(y_1, -\Delta X b | X_1, X_2, \alpha) dy_1 \right) \Delta X_j \Delta X_k \\
\leq H(X, \alpha)
$$

for some continuous and integrable function $H$ (this follows from Assumptions M.3 and E.5). This means that we can find the derivative of $G$ by differentiating under the expectation on the
right-hand side of (A2):

\[
\frac{\partial G_{2,k}(\beta)}{\partial \beta_j} = -E\left[ P(Y_1^* > 0, Y_2^* > 0 | X_1, X_2, \alpha) \right.
\]
\[
\times \left( 1\{\Delta X \beta > 0\}f_{Y_1^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right.
\]
\[
\times E\left[ Y_2^* | Y_1^* = \Delta X \beta, X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0 \right]
\]
\[
+ 1\{\Delta X \beta < 0\}f_{Y_1^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right]
\]
\[
\times E\left[ Y_1^* | Y_2^* = \Delta X \beta, X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0 \right]
\]
\[
\left. - P(Y_1^* > \Delta X \beta > - Y_2^* | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right) \Delta X_j \Delta X_k \]
\]
\[
= -P(Y_1^* > 0, Y_2^* > 0)
\]
\[
\times E\left[ \left( 1\{\Delta X \beta > 0\}f_{Y_1^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right.
\right.
\]
\[
\times E\left[ Y_2^* | Y_1^* = \Delta X \beta, X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0 \right]
\]
\[
+ 1\{\Delta X \beta < 0\}f_{Y_1^*}(\Delta X \beta | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right]
\]
\[
\times E\left[ Y_1^* | Y_2^* = \Delta X \beta, X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0 \right]
\]
\[
\left. - P(Y_1^* > \Delta X \beta > - Y_2^* | X_1, X_2, \alpha, Y_1^* > 0, Y_2^* > 0) \right) \Delta X_j \Delta X_k \right] \Delta X_j \Delta X_k \]
(ii) is trivially satisfied for \( l = 4 \) (use Lebesgue dominated convergence theorem to justify differentiation under expectation).

The results from the empirical processes literature, summarized in Pakes and Pollard (1989), can be used to verify (iii). First, notice that it suffices to prove that for \( l = 1, \ldots, 4, k = 1, \ldots, K, \)

\[
\sup_{\|b-\beta\|<\delta_n} \left| G_{k,\ln}(b) - G_{k,\ln}(b) - G_{k,\ln}(\beta) \right|
= \sup_{\|b-\beta\|<\delta_n} \left| (G_{k,\ln}(b) - G_{k,\ln}(b)) - (G_{k,\ln}(\beta) - G_{k,\ln}(\beta)) \right|
= n^{-1/2} \sup_{\|b-\beta\|<\delta_n} \left| \nu_n f_{k,\ln}(\cdot, b) - \nu_n f_{k,\ln}(\cdot, \beta) \right|
= o_p(n^{-1/2})
\]

where we have used the linear operator notation of Pakes and Pollard (1989), with \( \nu_n \) denoting the standardized empirical process \( \sqrt{n}(P_n - P) \).

It follows from Lemma 2.17 of Pakes and Pollard, that it is sufficient to demonstrate that (a) \( F_{k,\ln} \) is a Euclidean class with an envelope \( F_{k,\ln} \) for which \( \text{E}[F_{k,\ln}(Z)^2] < \infty \), and (b) \( \text{E}[f_{k,\ln}(Z, b)^2] \) is continuous at \( b = \beta \).

Clearly, \( F_{k,\ln} \) and \( F_{k,\ln} \) are bounded in absolute value by \( |\Delta X_k| \) and by \( \{|Y_{k}^*| + |Y_{k}^*| |\Delta X_k| \} \), respectively. These are squared integrable by Assumptions M.2 and M.3. It is also clear that \( \text{E}[f_{k,\ln}(Z, b)^2] \) (\( l = 1, \ldots, 4 \)) are continuous at \( b = \beta \) (see the expression for \( V_f \)). Therefore, it only remains to be shown that \( F_{k,\ln} \) are Euclidian.

To see that this is the case, note that for each \( f_{k,\ln} \in F_{k,\ln} \) (\( l = 1, 3; \ k = 1, \ldots, K \)), there is a partition of \( R^{K+3} \) into \( L \) regions, all of which are finite intersections of a finite number of halfplanes, such that on each of the regions, \( f_{k,\ln} \) is linear. We can, therefore, mimic the argument leading to Example 2.11 in Pakes and Pollard to show that \( F_{k,1} \) and \( F_{k,3} \) are Euclidian.

For \( l = 2 \), notice that \( f_{k,1} \) can be written as

\[
f_{k,1}(Z, b) = 1\{ -\Delta X_k Y_{k}^* < \Delta X_k \Delta X b \Delta X_k Y_{k}^* \} \]
\[
\times (\Delta X_k Y_{k}^* - \Delta X_k Y_{k}^* - \Delta X_k \Delta X b)
+ 1\{ -\Delta X_k Y_{k}^* > \Delta X_k \Delta X b \Delta X_k Y_{k}^* \} \]
\[
\times (\Delta X_k Y_{k}^* - \Delta X_k Y_{k}^* - \Delta X_k \Delta X b).
\]

Now consider this a function of \( \Delta X_k, \Delta X_k Y_{k}^*, \Delta X_k Y_{k}^*, \) and \( \Delta X_k \Delta X \). As a function of \( \Delta X_k Y_{k}^*, \Delta X_k Y_{k}^*, \) and \( \Delta X_k \Delta X \), there is a partition of \( R^{K+3} \) into at most \( L \) regions (for some finite \( L \) depending only on \( K \)), such that on each of the regions, \( f_{k} \) is linear. It therefore follows by the same argument as above, that \( F_{k,2} \) is Euclidian.

\( F_{k,4} \) is Euclidian by Lemma 2.13 of Pakes and Pollard.

Condition (iv) is a direct consequence of the Central Limit Theorem.

Part (v) follows from Assumption P.1 (for \( l = 1, 2 \)) and is trivial for \( l = 3, 4 \) (for which the parameter space is \( R^5 \)).

**Proof of Theorem 3:** Proof follows immediately from the discussion in Pakes and Pollard (1989, page 1043).

**REFERENCES**


