ROBUST ESTIMATION BASED ON GROUPED-ADJUSTED DATA
IN CENSORED REGRESSION MODELS

KAZUMITSU NAWATA

TECHNICAL REPORT NO. 15
APRIL 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANTS SES 82-08180 AND DMS 82-19748

ECONOMETRIC WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ROBUST ESTIMATION BASED ON GROUPED-ADJUSTED DATA
IN CENSORED REGRESSION MODELS

KAZUMITSU NAWATA

Technical Report No. 15
April 1986

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANTS SES 82-08180 and DMS 82-19748

ECONOMETRICS WORKSHOP
FOURTH FLOOR, ENCINA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1 Introduction

In recent economic studies, censored regression models in which the dependent variables cannot be negative are widely used. Such models were first suggested by Tobin [1958]. He analyzed household expenditures, which cannot be negative, on durable goods. The estimator he suggested is known as the Tobit Maximum Likelihood Estimator (Tobit MLE). Amemiya [1973] proved that the Tobit MLE is strongly consistent and asymptotically normal with the asymptotic variance-covariance matrix equal to \((E^{2} \log L/\partial \theta \partial \theta')^{-1}\) where \(\theta=(\beta', \sigma)\). However, unlike standard regression models, the Tobit MLE is inconsistent under heteroscedasticity (Arabmazar and Schmidt [1982], and also Brown and Moffitt [1983]) and nonnormality (Goldberger [1980], and also Arabmazar and Schmidt [1982]).

More recently, Powell [1981] has proposed a modified least absolute deviation (LAD) estimator which is strongly consistent under heteroscedasticity and nonnormality. His estimator minimizes

\[ \sum \max(0, x_j' \beta). \]

Powell [1983] proved that the estimator is asymptotically normal with the asymptotic variance-covariance matrix

\[ \frac{1}{4f(0)^2} \sum_{j \in \{k | x_j' \beta > 0\}} x_j x_j' \]

where \(f\) is the density function of the error term.

Paarsch [1983] conducted a Monte Carlo study in which he compared the Tobit MLE, Heckman's two-step estimator (Heckman [1976]), and Powell's LAD estimator under the Cauchy, Laplace and normal distributions. His findings are:
i) The Tobit MLE is better than the other two estimators for the Laplace and normal distributions.

ii) The Tobit MLE and Heckman's two-step estimator perform extremely poorly for the Cauchy distribution.

iii) Powell's LAD estimator performed reasonably well over all distributions, although it is a little worse than the Tobit MLE for the Laplace and normal distributions.

The major limitation of Powell's estimator is its computational difficulty. There are two problems in actual estimations. One is the differentiability of the minimand. Since the minimand is not differentiable at its minimum, we cannot use standard algorithms to calculate the estimator. The other is that the minimand has several local minima. So, it is very difficult to get the global minimum. Therefore, even if the model is very simple, it is very difficult and costly to estimate Powell's LAD estimator.

In this paper, I propose an alternative estimator, which is robust with respect to heteroscedasticity and nonnormality while still computationally feasible. The estimator consists of two different stages. In the first stage, the sample space of independent variables is divided into A cells. Then, Nawata's (1985) procedure is applied to those cells in which more than a$\frac{1}{4}$% (a$\frac{1}{4}$>50) of the observations have a positive dependent variable. Nawata's procedure is based on grouping and adjustment of the data. First, the proper domain of the independent variables is divided into a finite number of cells. The values of the adjusted dependent variable are calculated using the k-th round
estimator. The median of the adjusted dependent variable in each cell is estimated. Then, the next-round estimator of coefficients is calculated by OLS using the medians estimated in the previous stage. These operations are continued until the procedure converges. The second-stage estimator is calculated by the weighted Tobit MLE based on the data adjusted by the first-stage estimator. The estimator is defined in Chapter 2. Consistency of the estimator is proved in Chapter 3. The asymptotic distribution of the estimator is studied in Chapter 4, and the results of a Monte Carlo study are given in Chapter 5. The symbols used in this paper are summarized in Appendix A.
2. Model and Definitions

2.1 Model

The model studied in this paper is

\[
  y_j = \beta_0 + x_j^O \beta_1 + \tilde{u}_j = x_j' \beta + \tilde{u}_j, \quad j = 1, 2, \ldots, N,
\]

where

\[
  \tilde{u}_j = \begin{cases} 
  u_j & \text{if } u_j > -x_j' \beta \\
  -x_j' \beta & \text{otherwise.}
\end{cases}
\]

\(u_j\) are random variables with median 0. \(x_j^O\) are \(K\)-dimensional random vectors. Let \(\Omega\) be the union of the supports of the distribution functions of \(x_j^O\). \(\Omega\) satisfies the following assumptions. Additional assumptions on \(\{u_j\}\) and \(\{x_j^O\}\) will be stated whenever they are needed to prove subsequent theorems.

[Assumption 2.1]

i) \(\Omega\) is bounded and fixed.

ii) There is no perfect multicollinearity among elements of the vector \(x_j^O\).

iii) For any open neighborhood of \(z \in \Omega\), there exists \(\delta > 0\) such that \(\lim_{N \to \infty} P[|\xi|/N < \delta] = 0\) where \(\xi\) is the number of observations in the neighborhood of \(z\).

The sample estimator of the median of a sequence of real numbers \(\{a_j\} = \{a_1, a_2, \ldots, a_n\}\) is defined as
(2.2)  \[ M\{\{a_j\}\} = \begin{cases} a(n_1) & \text{if } n \text{ is odd} \\ (a(n_2) + a(n_3))/2 & \text{if } n \text{ is even} \end{cases} \]

where \( a(1) \leq a(2) \leq \ldots \leq a(n) \), \( n_1 = (n+1)/2 \), \( n_2 = n/2 \), and \( n_3 = (n+2)/2 \).

2.2 Definition of the Estimator

The definition of the estimator is given in this section. The definition of the estimator when \( x_j^0 \) are discrete and take finitely many values is given in 2.2.1. The general case is argued in 2.2.2.

2.2.1 Definition of the Estimator When the Independent Variables Are Discrete and Take Finitely Many Values

When \( x_j^0 \) are discrete and take finitely many values, \((x_1^0, x_2^0, \ldots, x_L^0)\), the estimator is defined by the following steps.

i) Calculate the median of the dependent variable \( Y_i \) for each \( x_i^0 \).

ii) Apply the weighted Tobit MLE.

The estimators \( \hat{\beta} \) and \( \hat{\sigma} \) maximize

\[
L(\alpha, \sigma) = \prod_i \Phi(-\sqrt{n_i} X_i' \alpha / \sigma) \prod_i \frac{\sqrt{n_i}}{\sqrt{2\pi}} e^{-n_i (Y_i - X_i' \alpha)^2 / 2 \sigma^2},
\]

where \( \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2} \, d\zeta \).

\( X_i' = (1, x_i^0) \),

\( I_i = \{ j : x_j^0 = x_i^0 \} \) (\( j \in I_i \) if and only if \( x_j^0 = x_i^0 \)),

\( Y_i = \) the median of the dependent variable for \( x_i^0 \).
\[ M\{y_j \mid j \in I_i\} \],

\[ m_i = \begin{cases} 1 & \text{if } y_i > 0 \\ 0 & \text{if } y_i = 0 \end{cases} \]

\[ n_i = \text{the number of elements in } I_i, \]

0 indicates \( m_i = 0 \) cells, and

1 indicates \( m_i = 1 \) cells.

\[ \]

2.2.2 Definition of the Estimator (General Case)

The estimation procedure consists of two different stages.

The first stage has the following steps.

i) Divide \( \Omega \) into \( A \) nonoverlapping cells \( \{\tilde{S}_\lambda\} \) which satisfy:

a) The distance between any two points in each cell goes to zero as the number of observations \( N \) goes to infinity.

b) \( A = o(\sqrt{N}) \).

c) \( \mu(\tilde{S}_\lambda) > b \mu(\Omega) / A, \lambda = 1, 2, \ldots, A \), for some \( b > 0 \) where \( \mu \) is the Lebesgue measure in \( \mathbb{R}^K \).

ii) Let \( S_{a_1} \) be the union of the cells in which more than 100\% of \( y_j \)'s are positive where \( 1 > a_1 > 0.5 \).

iii) Apply the following procedure (Nawata [1985]) to the observations which belong to \( S_{a_1} \). The procedure consists of the following steps.

a) Divide \( S_{a_1} \) into \( L^* + 1 \) nonoverlapping cells

\[ S^*_1, S^*_2, \ldots, S^*_{L^*}, S_{a_1} \] and take K-dimensional
vectors \( x_1^*, x_2^*, \ldots, x_L^* \) so that
\( \{S_i^*\} = \{S_1^*, S_2^*, \ldots, S_L^*\} \) and \( \{x_i^*\} = \{x_1^*, x_2^*, \ldots, x_L^*\} \)
satisfy the following conditions.

a.1) \( S_i^* \) are bounded.

a.2) \( \tilde{X}^* \tilde{X}^* \) converges to a nonsingular matrix,

where \( \tilde{X}_i^* = x_i^* - \sum_{v=1}^{L^*} x_v^*/L^*, \) and
\[ \tilde{X}^* = (\tilde{X}_1^*, \tilde{X}_2^*, \ldots, \tilde{X}_L^*) \] .

a.3) There exists \( \gamma < 1 \) such that
\[
\lim_{N \to \infty} P\left[ \max_{K} \sum_{L^*} a_{ij} \sum_{d_{ik} < \gamma} = 0 \right],
\]
where \( a_{ij} \) is the \((j,i)\)-th elements
of \( (\tilde{X}^* \tilde{X}^*)^{-1/2} \tilde{X}^* \); and
\[
d_{ik} = \sup_{\xi \in S_i^*} \text{the } k\text{-th element of } (\xi - x_i^*) .
\]

a.4) There exists \( \delta > 0 \) such that
\[
P[n_i^*/N < \delta] \to 0 \text{ for all } i, \text{ where } n_i^* \text{ is the number of observations in } S_i^*.
\]

As shown in Theorem 3.2, these cells and vectors exist under very general conditions. Note that if they do not exist for any \( a_i > 0.5 \), it is impossible to get a consistent estimator by this procedure. The first-stage estimator is estimated by the \( L^* \) observations which belong to \( U S_i^* \). Let the initial round estimator \( \beta_1^0 = 0 \).
b) Using the k-th round estimator $\hat{\beta}_1^k$, adjust the values of the dependent variable in each cell. The adjusted dependent variable $y_j^*(\hat{\beta}_1^k)$ is defined as

$$y_j^*(\hat{\beta}_1^k) = y_j^* - (x_j^0 - x_j^*)'\beta$$

where $x_j^0 \in S_1^*$.

c) Estimate the median of the adjusted dependent variable of the observations which belong to $S_{a_1}$ in each cell.

d) Using (c), calculate the (k+1)-th round estimator $\hat{\beta}_1^{k+1} = T(\hat{\beta}_1^k)$ by ordinary least squares.

e) Repeat (b) through (d).

f) Define the first-stage estimator $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \begin{cases} 
\lim_{k \to \infty} \hat{\beta}_1^k & \text{if } T \text{ is a contraction mapping} \\
0 & \text{otherwise.} 
\end{cases}$$

The purpose of this stage is to get a consistent estimator of order $N^{-1/2}$. As I show in Chapter 4, the asymptotic properties of the second-stage estimator do not depend on the first-stage estimator if it is a consistent estimator of order $N^{-1/2}$.

The second-stage estimator $\hat{\beta}$ is then calculated by the weighted Tobit MLE using the data which has been adjusted by the first-stage estimator. Note that the observations which do not belong to $S_{a_1}$ are also used in this stage. The procedure consists of the following steps.

1) Divide the proper subset of the space of $x_j^0$ into $L$ nonoverlapping, nonempty, convex cells $S_1, S_2, \ldots, S_L$.

2) Calculate the median of the adjusted dependent variable in
each cell.

iii) Apply the weighted Tobit MLE.

Let

\[ I_i = \text{the index set of the } i\text{-th cell and } j \in I_i \text{ if and only if } x_j \in S_i, \]

\[ n_i = \text{the number of elements in } I_i, \]

\[ x_i^0 = \text{the mean of the } x_j \text{ in the } i\text{-th cell} \]

\[ = \frac{1}{n_i} \sum_{j \in I_i} x_j, \]

\[ x_i = (1, x_i^0'), \]

\[ \bar{x}_i^0 = \text{the } "\text{minimum } \text{value of } x_j \text{ in the } i\text{-th cell which attains } \min_{k \in I_i} x_k^0, \]

\[ \bar{x}_i = (1, \bar{x}_i^0'), \]

\[ y_j(\bar{\theta}_1) = \text{the adjusted dependent variable} \]

\[ = y_j - (x_j^0 - x_i^0)'\bar{\theta}_1 \text{ for } j \in I_i, \]

\[ y_i = \text{the median of adjusted dependent variables in the } i\text{-th cell,} \]

\[ = M\{y_j(\bar{\theta}_1) : j \in I_i\} = X_i' \hat{\beta} + \tilde{\varepsilon}_i \text{ where } M \text{ is a sample estimator of the median defined in (2.2),} \]

and

\[ M_i^* = \text{the maximum value of } y_j(\bar{\theta}_1) \text{ which satisfies } y_j(\bar{\theta}_1) < y_i \text{ in the } i\text{-th cell,} \]

\[ = \max_{j \in I_i} \{y_j(\bar{\theta}_1) : y_j(\bar{\theta}_1) < y_i\}. \]

The median is labeled as "observed" in the i-th cell, if
\[(2.5) \quad I_i \cap \{j: y_j = 0\} \cap \{j: y_j(\bar{\beta}_1) \leq y^*_j\} = \emptyset.\]

Let \(\{y^*_j\}\) be uncensored dependent variables. If \((2.5)\) is satisfied, \(M(\{y^*_j(\bar{\beta}_1): j \in I_1\}) = M(\{y^*_j(\bar{\beta}_1): j \in I_1\})\) where \(y^*_j(\bar{\beta}_1) = y^*_j - (x_j^0 - X_i^0)'\bar{\beta}_1\) for \(j \in I_1\). Therefore, we can actually observe the uncensored median of the adjusted dependent variable in this case. Define,

\[
m_i = \begin{cases} 
1 & \text{if the median is observed} \\
0 & \text{otherwise.}
\end{cases}
\]

Then the second-stage estimator \(\hat{\beta}\) maximizes:

\[
(2.6) \quad L(\alpha, \sigma) = \prod_{0} \Phi(-\sqrt{n_i} X_i'\alpha/\sigma) \prod_{1} \frac{\sqrt{n_i}}{\sqrt{2\pi\sigma}} e^{-\frac{n_i}{2}(Y_i - X_i'\alpha)^2/2\sigma^2},
\]

where \(0\) indicates \(m_i = 0\) cells, and \(1\) indicates \(m_i = 1\) cells.

Note that the value of the independent variables which represents each cell must be the mean if the median is "observed" and the "minimum" value if the median is not "observed".
3. **Consistency of the Estimator**

Consistency of \( \hat{\beta} \) when the independent variables are discrete and take finitely many values is proved in 3.1. The general case is argued in 3.2.

3.1 **Consistency of the Estimator When the Independent Variables Are Discrete and Take Finitely Many Values**

The following theorem proves the consistency of the estimator when the independent variables take finitely many values.

[Theorem 3.1]

Let \( x_j^0 \) be discrete and take a finite number of different values. The estimator, \( \hat{\beta} \), which maximizes (2.3) is consistent under the following conditions.

i) There exists \( \delta > 0 \) such that \( f_j(u) > \delta \) in the neighborhood of zero for all \( j \) where \( f_j \) is the density function of \( u_j \).

ii) \( \operatorname{plim}_{N \to \infty} n_1/N = p_1 > 0 \) for all \( i \).

iii) \( \sum_{1} (n_1/N) X_i X'_i \) converges to a positive definite matrix \( A \) in probability.

iv) \( X'_i \beta \neq 0 \) for all \( i \).

v) \( \operatorname{plim}_{N \to \infty} M[\{u_j: j \in I_1 \}] = 0 \) for all \( i \).

[Proof]

Let \( \hat{\beta}^* \) and \( \hat{\sigma}^* \) maximize
(3.1) \[ L_1(\alpha, \sigma) = \prod_{1}^{\sqrt{n_1}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{n_1(x' - X'\alpha)^2}{2\sigma^2}}. \]

Then

(3.2) \[ \hat{\beta}^* = (\sum_{1}^{\sqrt{n_1}} n_i x_i x_i')^{-1} \sum_{1}^{\sqrt{n_1}} n_i x_i y_i \]

\[ = \beta + (\sum_{1}^{\sqrt{n_1}} n_i x_i x_i')^{-1} \sum_{1}^{\sqrt{n_1}} n_i x_i \tilde{\varepsilon}_i. \]

If \( m_1 = 1, \tilde{\varepsilon}_1 = \varepsilon_1 \). Therefore,

(3.3) \[ \hat{\beta}^* = \beta + \left[ \sum_{1}^{\sqrt{n_1}} (n_i/N) x_i x_i' \right]^{-1} \sum_{1}^{\sqrt{n_1}} (n_i/N) x_i \varepsilon_i. \]

Since \( \sum_{1}^{\sqrt{n_1}} (n_i/N) x_i x_i' \rightarrow A \) and \( \varepsilon_i \rightarrow 0 \),

(3.4) \[ \hat{\beta}^* \rightarrow \beta. \]

From (3.1),

(3.5) \[ \hat{\sigma}^2 = \sum_{1}^{\sqrt{n_1}} n_1(x_i - X_i' \hat{\beta}^*)/L_1. \]

Since \( \text{plim}_{N \rightarrow \infty} n_k/N = p_k > 0 \) for all \( k \),

(3.6) \[ \frac{\hat{\sigma}^2}{n_k} = \sum_{1}^{\sqrt{n_1}} n_1 [X_i' (\beta - \hat{\beta}^*) + \varepsilon_i] \rightarrow 0. \]

From (3.5) and (3.6),

(3.7) \[ L_1(\hat{\beta}^*, \hat{\sigma}^2) \rightarrow \infty. \]

Let

(3.8) \[ \beta^* = \beta + \eta. \]

Then \(-\sqrt{n_1} X_i' \hat{\beta}^*/\hat{\sigma}^*) = \left(\sqrt{n_1} X_i' \beta/\hat{\sigma}^*)\left(1 + n_1/X_i' \beta\right)\right). \) Since \( \eta \rightarrow 0 \) and \( X_i' \beta \neq 0 \),
$1 + \eta / X' \beta + 1$ in probability. Hence,

\[(3.9) \quad -\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast} \xrightarrow{\text{P}} \begin{cases} \infty & \text{if } X'_{i} \beta < 0 \\ -\infty & \text{if } X'_{i} \beta > 0. \end{cases} \]

Therefore,

\[(3.10) \quad \Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) \xrightarrow{\text{P}} \begin{cases} 1 & \text{if } X'_{i} \beta < 0 \\ 0 & \text{if } X'_{i} \beta > 0. \end{cases} \]

Here, for any $\varepsilon > 0$ in $(0,1)$,

\[
\{ \Pi \Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < 1 - \varepsilon \} \subseteq \bigcup_{i=1}^{L} \{ \Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < (1-\varepsilon)^{1/L}, \varepsilon_{i} \leq X'_{i} \beta \}. 
\]

Hence,

\[(3.11) \quad P[\Pi \Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < 1 - \varepsilon] \leq \sum_{i=1}^{L} P[\Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < (1-\varepsilon)^{1/L}, \varepsilon_{i} \leq X'_{i} \beta]. 
\]

Here,

\[P[\Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < (1-\varepsilon)^{1/L}] \to 0 \text{ if } X'_{i} \beta < 0 , \]

and

\[P[\varepsilon_{i} \leq X'_{i} \beta] \to 0 \quad \text{if } X'_{i} \beta > 0 . \]

Therefore,

\[(3.12) \quad P[\Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) < (1-\varepsilon)^{1/L}, \varepsilon_{i} \leq X'_{i} \beta] \to 0 \text{ for all } i . \]

From (3.11) and (3.12),

\[(3.13) \quad \Pi \Phi(-\sqrt{n} X'_{i} \hat{\beta}^{\ast} / \hat{\sigma}^{\ast}) \xrightarrow{\text{P}} 1. \]
From (3.7) and (3.13),

\[ L(\hat{\beta}, \hat{\sigma}) \overset{P}{\to} \omega. \]  

Here,

\[ \sum_{l} (n_{l}/N)(Y_{l}-X_{l}^{T}\alpha)^2 = \sum_{l} (n_{l}/N)[X_{l}^{T}(\beta-\alpha)+\varepsilon_{l}]^2. \]  

Let \( \Theta_{\varepsilon} = \{ \alpha: \|\alpha-\beta\|^{2} > \varepsilon \} \) where \( \|\alpha-\beta\|^{2} = (\alpha-\beta)^{T}(\alpha-\beta) \). Since \( \varepsilon \to 0 \),

\[ \inf_{\alpha \in \Theta_{\varepsilon}} \sum_{l} (n_{l}/N)[X_{l}^{T}(\beta-\alpha)+\varepsilon_{l}]^2 \overset{P}{\to} \inf_{\alpha \in \Theta_{\varepsilon}} \sum_{l} (n_{l}/N)[X_{l}^{T}(\beta-\alpha)]^2. \]  

Let \( \lambda^{*}_{S} \) be the smallest characteristic root of \( \sum_{l} (n_{l}/N)X_{l}^{T}X_{l} \). Then

\[ \inf_{\alpha \in \Theta_{\varepsilon}} \sum_{l} (n_{l}/N)[X_{l}^{T}(\alpha-\beta)]^2 > \lambda^{*}_{S} \varepsilon. \]  

From (3.1), \( \lambda^{*}_{S} \) converges to \( \lambda_{S} > 0 \) in probability where \( \lambda_{S} \) is the smallest characteristic root of \( A \).

From (3.15)-(3.17),

\[ P\left( \inf_{\alpha \in \Theta_{\varepsilon}} \sum_{l} (n_{l}/N)(Y_{l}-X_{l}^{T}\alpha)^2/N < \delta/2 \right) \to 0, \]  

where \( \delta = \lambda S \varepsilon \). If \( \sum_{l} (n_{l}/N)(Y_{l}-X_{l}^{T}\alpha)^2 > \delta/2 \),

\[ L(\alpha, \sigma) = \prod_{0}^{\sqrt{n_{l}/X_{l}^{T}\alpha}/\sigma} \prod_{1}^{\sqrt{2\pi}\sigma} e^{-n_{l}/(Y_{l}-X_{l}^{T}\alpha)^2}/2\sigma^2} \]

\[ \leq \left( \sqrt{n}/\sqrt{2\pi}\sigma \right) \frac{L}{1} e^{-\delta N/4\sigma^2}. \]
Therefore, there exists $\Delta < \infty$ such that,

$$\sup_{\alpha \in \Theta_\epsilon, \sigma} L(\alpha, \sigma) < \Delta \quad \text{if} \quad \inf_{\alpha \in \Theta_\epsilon} \sum_{i=1}^{n_i} \frac{(n_i/N)(Y_i - X_1')\alpha)^2}{\alpha \epsilon} > \Delta/2. \quad (3.20)$$

From (3.20),

$$\mathbb{P}[\sup_{\alpha \in \Theta_\epsilon, \sigma} L(\alpha, \sigma) > \Delta] < \mathbb{P}\left[ \inf_{\alpha \in \Theta_\epsilon} \sum_{i=1}^{n_i} \frac{(n_i/N)(Y_i - X_1')\alpha)^2}{\alpha \epsilon} < \frac{\Delta}{2}\right]. \quad (3.21)$$

From (3.18) and (3.21),

$$\mathbb{P}\left[ \sup_{\alpha \in \Theta_\epsilon, \sigma} L(\alpha, \sigma) > \Delta \right] + 0. \quad (3.22)$$

Here,

$$\mathbb{P}[\rho(\beta, \hat{\beta}) > \epsilon] = \mathbb{P}[L(\hat{\beta}, \hat{\sigma}) > \Delta, \rho(\beta, \hat{\beta}) > \epsilon]$$

$$+ \mathbb{P}[L(\hat{\beta}, \hat{\sigma}) < \Delta, \rho(\beta, \hat{\beta}) > \epsilon], \quad (3.23)$$

and

$$\mathbb{P}[L(\hat{\beta}, \hat{\sigma}) > \Delta, \rho(\beta, \hat{\beta}) > \epsilon] < \mathbb{P}\left[ \sup_{\alpha \in \Theta_\epsilon, \sigma} L(\alpha, \sigma) > \Delta\right]. \quad (3.24)$$

Since $\hat{\beta}$ and $\hat{\sigma}$ maximize $L$,

$$\mathbb{P}[L(\hat{\beta}, \hat{\sigma}) < \Delta, \rho(\beta, \hat{\beta}) > \epsilon] < \mathbb{P}[L(\hat{\beta}, \hat{\sigma}) < \Delta] < \mathbb{P}[L(\hat{\beta}^*, \hat{\sigma}^*) < \Delta]. \quad (3.25)$$

From (3.22) and (3.24),

$$\mathbb{P}[L(\hat{\beta}, \hat{\sigma}) > \Delta, \rho(\beta, \hat{\beta}) > \epsilon] + 0. \quad (3.26)$$

From (3.14) and (3.25),

$$\mathbb{P}[L(\hat{\beta}, \hat{\sigma}) < \Delta, \rho(\beta, \hat{\beta}) > \epsilon] + 0. \quad (3.27)$$

From (3.23), (3.26), and (3.27),

$$\mathbb{P}[\rho(\beta, \hat{\beta}) > \epsilon] \rightarrow 0 \text{ for any } \epsilon > 0. \quad (3.28)$$

Hence,

$$\hat{\beta} \rightarrow \beta. \quad (3.29)$$

QED.
3.2 Consistency of the Estimator (General Case)

Consistency of the estimator in the general case is proved in this section. Consistency of $\hat{\beta}_1$ is proved in Theorem 3.2. The proof of the consistency of $\hat{\beta}$ is in Theorem 3.3. The following proposition is used in the proof.

[Proposition 3.1]
Let $e_j = v_j w + u_j$, and $\tilde{e}_j = v_j w + \tilde{u}_j$, $j = 1, 2, \ldots, n$,
where $u_j, \tilde{u}_j, v_j, w$ and $n$ are random variables which satisfy the following conditions.

i) $v_j$ are bounded.

ii) $\lim_{N \to \infty} \frac{n}{N} = p > 0$.

iii) $\lim_{N \to \infty} w = 0$.

iv) $\lim_{N \to \infty} \mu\{u_j\} = 0$.

v) There exists $\delta > 0$, such that $f_j(0) > \delta > 0$ in the neighborhood of zero for all $j$, where $f_j$ is the density function of $u_j$.

vi) There exists $a > 0$ such that $\tilde{u}_j = u_j$ if $u_j > -a$.

Then

\[(3.30)\quad P\{M\{e_j\} \neq M\{\tilde{e}_j\}\} \to 0.\]

Note that (3.30) implies $P\{M\{u_j\} \neq M\{\tilde{u}_j\}\} \to 0$. 
[Proof]

Let \( e_1 > e_2 > \ldots > e_n \), and

\[
(3.31) \quad n^* = \begin{cases} 
  (n+1)/2 & \text{if } n \text{ is odd} \\
  (n/2)+1 & \text{if } n \text{ is even},
\end{cases}
\]

where \( n \) is the number of the elements in the set.

Since \( \tilde{u}_j = u_j \) if \( u_j > -a \), \( \tilde{e}_j = e_j \) if \( e_j > -a+v^* \), where \( v^* = \sup_{v_j} v^\prime v \).

If \( e_{(n^*)} > -a+v^* \), then \( e_j = \tilde{e}_j \) for \( 1 \leq j < n^* \). Therefore,

\[
(3.32) \quad M[\{e_j\}] = M[\{\tilde{e}_j\}] \quad \text{if} \quad e_{(n^*)} > -a+v^* .
\]

Consequently,

\[
(3.33) \quad P(M[\{e_j\}] = M[\{\tilde{e}_j\}]) > P(e_{(n^*)} > -a+v^*) .
\]

From assumptions (iv) and (v), \( e_{(n^*)} \neq M[\{e_j\}] \) in probability. Therefore,

\[
(3.34) \quad P(|e_{(n^*)} - M[\{e_j\}]| > \frac{a}{2}) \to 0 .
\]

If \( |e_{(n^*)} - M[\{e_j\}]| < a/2 \),

\[
(3.35) \quad P(e_{(n^*)} > -a+v^*) > P(M[\{e_j\}] > \frac{a}{2} + v^* ) .
\]

Here

\[
(3.36) \quad P(M[\{e_j\}] > \frac{a}{2} + v^*) = 1 - P(M[\{e_j\}] \leq \frac{a}{2} + v^*) \]

\[
> 1 - P(M[\{e_j\}] \leq \frac{a}{4}, v^* \leq \frac{a}{4}) - P(v^* > \frac{a}{4} ) .
\]
Since \(|e_j - u_j| < \|v\*\|, \|M[{e_j}] - M[{u_j}]\| < \|v\*\|\). Since \(v\*=0\) in probability, \(P(M[{e_j}] < a/4) \rightarrow 0\) and \(P(v\*a/4) \rightarrow 0\). Hence,

\[(3.37) \quad P(M[{e_j}] > \frac{a}{2} + v\*) \rightarrow 1.
\]

From (3.33),(3.34),(3.35), and (3.37),

\[(3.38) \quad P(M[{e_j}] \notin M[{\tilde{e}_j}]) \rightarrow 0.
\]

QED.

Next I prove that the first-stage estimator \(\tilde{\beta}_1\) is consistent.

[Theorem 3.2]

The first-stage estimator \(\tilde{\beta}_1\) is consistent under the following assumptions.

i) For given \(a_1 > 0.5\), there exist \(a^* > a_1\) and \(\delta_1 > 0\) such that:

a) \(\text{plim}_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}(a^*) x_j x_j'\) exists and is positive definite, where

\[\mathbb{1}(a^*) = \begin{cases} 
1 & \text{if } F_j(x_j^i) > a^* \\
0 & \text{otherwise}, 
\end{cases}\]

and

\(F_j\) is the distribution function of \(u_j\).

b) If \(F_k(x_k^i) > a^*\), \(F_j(x_j^i) > a_1\) for any \(x_j^\circ\) which satisfies \(\|x_j^\circ - x_k^\circ\| < \delta_1\).

ii) There exist \(\delta_2 > 0\) and \(\delta_3 < \infty\) such that \(\delta_2 < f_j(u) < \delta_3\) in the
neighborhood of zero, where $f_j$ is the density function of $u_j$.

iii) \( \lim_{N \to \infty} M(\{u_j : j \in I^*_i\}) = 0 \) for all \( i \), where \( I^*_i \) is an index set of \( S^*_i \) (\( j \in I^*_i \) if and only if \( x_j^o \in S^*_i \)).

iv) There exist \( \delta_4 > 0 \) and \( N_0 \) such that

\[
\min_{1 \leq \lambda < \Lambda} \frac{\sum_{j=1}^{N} P[x_j^o \in \tilde{S}_\lambda]}{(N/\Lambda)} > \delta_4 \quad \text{if} \quad N > N_0 .
\]

v) \( (x_j^o, F_j(x_j^o), u_j) \) and \( (x_k^o, F_k(x_k^o), u_k) \) are independent if \( j \neq k \).

Note that assumption (iv) is satisfied if all distribution functions of \( x_j^o \) have the common support.

[Proof]

Let \( \tilde{S}_\lambda \) be the \( \lambda \)-th cell. Then

\[
(3.39) \quad S = \bigcup_{\lambda} S^*_\lambda
\]

where \( S^*_\lambda \) is a cell where \( n_\lambda^+/n_\lambda > a_1 \),

\( n_\lambda \) is the number of observations in the \( \lambda \)-th cell, and \( n_\lambda^+ \) is the number of observations such that \( y_j > 0 \) in the \( \lambda \)-th cell.

Define

\[
(3.40) \quad \phi_j^\lambda = \begin{cases} 1 & \text{if } x_j^o \in \tilde{S}_\lambda \\ 0 & \text{otherwise} \end{cases}
\]
\( (3.41) \quad \psi_j^\lambda = (\phi_j^\lambda - \Pr[x_j^\circ \in S_\lambda]) \phi_j^\lambda, \)

\( (3.42) \quad \xi_j = \begin{cases} 
1 & \text{if } y_j > 0 \\
0 & \text{otherwise,} 
\end{cases} \)

\( (3.43) \quad \xi_j^\lambda = \xi_j \phi_j^\lambda, \)

and

\( (3.44) \quad \eta_j^\lambda = [\xi_j^\lambda - F_j(x_j^\lambda)] \phi_j^\lambda. \)

Then

\( (3.45) \quad E \phi_j^\lambda = 0, \)

\( (3.46) \quad \text{Var}(\psi_j) = (1 - \Pr[x_j^\circ \in S_\lambda]) \Pr[x_j^\circ \in \tilde{S}_\lambda] \prec \Pr[x_j^\circ \in S_\lambda], \)

\( (3.47) \quad E \eta_j^\lambda = 0, \)

and

\( (3.48) \quad \text{Var}(\eta_j^\lambda) = \int_{\tilde{S}_\lambda} F(x_j^\lambda)[1 - F(x_j^\lambda)] d\Pr(x_j^\circ) \prec \Pr[x_j^\circ \in \tilde{S}_\lambda]/4. \)

Here,

\( (3.49) \quad \Pr\left[ \left| \sum_{j=1}^{N} \eta_j^\lambda / n_\lambda \right| > \varepsilon \right] < \Pr[n_\lambda < \bar{n}_\lambda, \left| \sum_{j=1}^{N} \eta_j^\lambda / \bar{n}_\lambda \right| > \varepsilon] + \Pr[n_\lambda < \bar{n}_\lambda] \)

\[ < \Pr\left[ \left| \sum_{j=1}^{N} \eta_j^\lambda / \bar{n}_\lambda \right| > \varepsilon \right] + \Pr[n_\lambda < \bar{n}_\lambda]. \]
Here \( E n_\lambda = \sum_{j=1}^{N} P[x_j^0 \in \tilde{S}_\lambda] \). By Chebyshev's inequality,

\[
(3.50) \quad P\left[ \left| \sum_{j=1}^{N} \frac{\phi_j^\lambda}{E n_\lambda} \right| > \zeta \right] < \sum_{j=1}^{N} \frac{\text{Var}(\phi_j^\lambda)}{(\zeta E n_\lambda)^2} = 1/(\zeta^2 E n_\lambda) .
\]

Since \( \sum_{j=1}^{N} \phi_j^\lambda = n_\lambda - E n_\lambda \),

\[
(3.51) \quad P[n_\lambda/E n_\lambda < 1-\zeta] < P\left[ \left| (n_\lambda - E n_\lambda)/E n_\lambda \right| > \zeta \right] < 1/(\zeta^2 E n_\lambda) .
\]

Let \( \zeta = 1/2 \). Then (3.51) gives,

\[
(3.52) \quad P[n_\lambda < E n_\lambda/2] < 6/E n_\lambda .
\]

By Chebyshev's inequality, for any \( \varepsilon > 0 \),

\[
(3.53) \quad P\left[ \left| \sum_{j=1}^{N} \frac{n_j^\lambda/n_\lambda}{E n_\lambda} \right| > \varepsilon \right] < 1/(4\varepsilon^2 E n_\lambda) .
\]

Take \( \bar{n}_\lambda = E n_\lambda/2 \). From (3.49), (3.52), and (3.53),

\[
(3.54) \quad P\left[ \left| \sum_{j=1}^{N} \frac{n_j^\lambda/n_\lambda}{E n_\lambda} \right| > \varepsilon \right] < (4 + 1/\varepsilon^2)/E n_\lambda .
\]

From assumption ii), there exist \( a_1^* \), \( a_2^* \), and \( \delta \) such that,

a) \( 0.5 < a_1^* < a_1 \), \( 0.5 < a_2^* < a_2 \), and \( \delta > 0 \), and

b) if \( F_j(x_j^\beta) < a_1^* \), \( F_j(x_j^\beta) < a_2^* \) for any \( x_j^0 \) which satisfies \( \| x_j^0 - x_k^0 \| < \delta \).

Make the size of each cell small enough so that \( |(z_1^0 - z_2^0)'\beta| < \delta \) for all \( z_1^0, z_2^0 \in S_\lambda \) where \( z_1^1 = (1, z_1^0) \). Suppose \( x_k^0 \in \tilde{S}_\lambda \) and

\[
P_k(x_k^\beta) < a_k^* \text{ for some } k . \quad \text{Then, } P_j(x_j^\beta) < s_j^* \text{ for any } j \in I_\lambda .
\]

Therefore, for this cell,
(3.55) \[ P\left[ \sum_{j=1}^{N} \frac{\lambda_j}{n_{\lambda}} > \varepsilon \right] > P\left[ \sum_{j=1}^{N} \frac{\varepsilon_j}{n_{\lambda}} > \sum_{j=1}^{N} \phi_j f_j(x_j^j) / n_{\lambda} + \varepsilon \right] > P[n_{\lambda}^+ / n_{\lambda} > a_2 + \varepsilon] . \]

Take \( \varepsilon = (a_1 - a_2) / 2 \). Let \( A_{\lambda} = U \{ x_j^o \in S_j \lambda \text{ and } f_j(x_j^j) < a_1 \} \) and \( B_{\lambda} = \{ n_{\lambda}^+ / n_{\lambda} > a_1 \} \). Then from (3.54) and (3.55),

(3.56) \[ P[B_{\lambda} | A_{\lambda}] < P[n_{\lambda} / n > (a_1 + a_2) / 2 | A_{\lambda}] < c_1 / E n_{\lambda} . \]

where \( c_1 = 1 / [(a_1 - a_2)^2 + 4] \).

Hence,

(3.57) \[ P[A_{\lambda} \cap B_{\lambda}] < c_1 / E n_{\lambda} . \]

Here,

(3.58) \[ P[\{ S_{a_1} \text{ includes } x_j^o \text{ which satisfies } f_j(x_j^j) < a_1 \}] \equiv P[ \bigcup_{j=1}^{N} \{ x_j^o \in S_{a_1} \text{ and } f_j(x_j^j) < a_1 \} ] = P[ \bigcup_{j=1}^{N} \{ A_{\lambda} \cap B_{\lambda} \} ] = \sum_{\lambda=1}^{A} P[ A_{\lambda} \cap B_{\lambda} ] \leq \sum_{\lambda=1}^{A} c_1 / E n_{\lambda} . \]

Since \( A = o(\sqrt{N}) \), assumption (iv) gives \( 1 / E n_{\lambda} = o(\Delta) \) for all \( \lambda \).

Therefore, (3.58) gives
(3.59) \[ P\{S_{a_1} \text{ includes } x_j^0 \text{ which satisfies } F_j(x_j' \beta) < a^*_1\} \to 0. \]

In the same way,

(3.60) \[ P\{S_{a_1} \text{ includes all } x_j^0 \text{ which satisfy } F_j(x_j' \beta) > a^*_1\} \to 1. \]

From assumption (i), (3.60), and Theorem 2.1 of Nawata [1985], there exist cells \( \{S_i^*\} \) and \( K \)-dimensional vectors \( \{X_i^*\} \) which satisfy the condition (a.1) through (a.4) in section 2.2.2.

From assumption (ii), there exists \( \delta^* > 0 \) such that \( F_j^{-1}(a_{1i}^*) > \delta^* \) for all \( j \). Hence, from (3.59),

(3.61) \[ P\{\text{all } x_j^0 \text{ in } S_{a_1} \text{ satisfy } x_j' \beta > \delta^*\} \to 1. \]

Here, \( \tilde{u}_j = u_j \) if \( u_j > x_j' \beta \). Therefore, from (3.61) and Proposition 3.1,

(3.62) \[ \lim_{N \to \infty} M[\tilde{u}_j : j \in I_i^*] = 0. \]

Therefore, by the same argument of Theorem 3.4 of Nawata [1985], the first-stage estimator given by \( \{S_i^*\} \) and \( \{X_i^*\} \) is consistent.

\[ \text{QED.} \]

Next I prove that the second-stage estimator is consistent.

[Theorem 3.3]

The second stage estimator \( \hat{\beta} \) defined in (2.6) is consistent under the assumptions in Theorem 3.3 and the following assumptions.

1) \( S_i \) are fixed cells.
ii) $b_i = \liminf_{j \in I_1} x_j^i \beta \neq 0$ for all $i$.

iii) If $b_i < 0$, there exists $\delta > 0$ such that the number of observations which which satisfy $x_j^i \beta < -\delta$ in the $i$-th cell goes to infinity in probability as $N$ goes to infinity.

iv) $\lim_{N \to \infty} n_i/N = p_i > 0$ for all $i$.

v) $\sum_{1}^N (n_i/N) X_i' X_i$ converges to a positive definite matrix $A$ in probability.

vi) $\lim_{N \to \infty} M_1[\{u_j: j \in I_1\}] = 0$ for all $i$.

[Proof]

First I show,

(3.63) \quad P[m_i = 0 | b_i > 0] \to 0,

and

(3.64) \quad P[m_i = 1 | b_i < 0] \to 0.

Here, for $j \in I_1$,

(3.65) \quad y_j(\beta) = y_j - (x_j^i - x_j^0)'(\beta - \beta_1)

\quad \quad \quad = x_j^i \beta + (x_j^i - x_j^0)'(\beta - \beta) + \tilde{u}_j

\quad \quad \quad = x_j^i \beta + \tilde{e}_j,

where $\tilde{e}_j = (x_j^i - x_j^0)'(\beta - \beta_1) + \tilde{u}_j$. 
From Proposition 3.1, 

\begin{equation}
(3.66) \quad P \left[ \left\{ \tilde{e}_j : j \varepsilon I_1 \right\} \right] \rightarrow M[\left\{ e_j : j \varepsilon I_1 \right\}] \quad \text{if } b_1 > 0,
\end{equation}

where \( e_j = (x_1^0 - x_j^0) \left( \bar{\beta}_1 - \beta_1 \right) + u_j \).

Here, 

\begin{equation}
(3.67) \quad P \left| M[\left\{ e_j : j \varepsilon I_1 \right\}] \right| \leq |M[\left\{ u_j : j \varepsilon I_1 \right\}]| + \sup_{j \in I_1} |x_1^0 - x_j^0| \left| \bar{\beta}_1 - \beta_1 \right|.
\end{equation}

Since \( \bar{\beta}_1 = \beta_1 \), 

\begin{equation}
(3.68) \quad P M[\left\{ e_j : j \varepsilon I_1 \right\}] \rightarrow 0.
\end{equation}

Let \( y_1(\bar{\beta}_1) > y_2(\bar{\beta}_1) > \ldots > y_{n_1}(\bar{\beta}_1) \),

and

\begin{align*}
n_i^* = \begin{cases} 
(n_i + 1)/2 & \text{if } n_i \text{ is odd} \\
n_i/2 + 1 & \text{if } n_i \text{ is even}.
\end{cases}
\end{align*}

Then, 

\begin{equation}
(3.69) \quad P \tilde{e}_{n_i^*} + M[\left\{ \tilde{e}_j : j \varepsilon I_1 \right\}].
\end{equation}

From (3.66), (3.68), and (3.69),

\begin{equation}
(3.70) \quad P \left[ e_{n_i^*} < -b_1/4 \left| b_1 > 0 \right. \right] \rightarrow 0.
\end{equation}

Here, \( m_i = 1 \) if

\begin{equation}
(3.71) \quad \sup_{j \in I_1} (x_1^0 - x_j^0) (\bar{\beta}_1 - y_{n_i^*}(\bar{\beta}_1)).
\end{equation}
Since \( y_{n_1^*}(\tilde{\beta}_1) = X'_{n_1} \hat{\beta} + e_{n_1^*} \), (3.71) is equivalent to

\[
(3.72) \quad X_i'(\tilde{\beta}_1 - \beta_1) - \inf_{j \in I_i} [x_j' \beta + x_j'(\tilde{\beta}_1 - \beta_1)] < e_{n_1^*}.
\]

Since \( \tilde{\beta}_1 \to \beta_1 \) and \( \inf_{j \in I_i} x_j' \beta \to b_1 \),

\[
(3.73) \quad P(X_i'(\tilde{\beta}_1 - \beta_1) - \inf_{j \in I_i} [x_j' \beta + x_j'(\tilde{\beta}_1 - \beta_1)] > -b_1/2) \to 0.
\]

From (3.70), (3.71), and (3.73), \( P[m_1 = 0 | b_1 > 0] \to 0 \).

Suppose \( b_1 < 0 \). From assumption (iii), there exists \( \delta > 0 \) such that the number of observations which satisfy \( x_j' \beta < -\delta \) in the \( i \)-th cell goes to infinity in probability. Since \( P[y_j = 0 | x_j' \beta < -\delta] > 1/2 \),

\[
(3.74) \quad P\{\text{there exists } j \text{ such that } y_j = 0, j \in I_i, \text{ and } x_j' \beta < -\delta\} \to 1.
\]

Let \( \kappa \) be the observation such that \( \kappa \in I_i \), \( x_{\kappa}' \beta < -\delta \) and \( y_\kappa = 0 \). Then

\[
(3.75) \quad y_\kappa(\tilde{\beta}_1) = -(x_{\kappa}' - x_i')'(\tilde{\beta}_1)
\]

\[= X_i' \beta - x_{\kappa}' \beta + (x_{\kappa}' - x_i')'(\tilde{\beta}_1 - \beta_1).
\]

Since \( \tilde{\beta}_1 \to \beta_1 \) in probability,

\[
(3.76) \quad P[y_\kappa(\tilde{\beta}_1) < X_i' \beta + \delta/2] \to 0.
\]

Suppose \( y_\kappa(\tilde{\beta}_1) > \delta/2 + X_i' \beta \). To get \( m_i = 1 \),

\[
(3.77) \quad e_{n_1^*} \geq \delta/2.
\]

Here,
(3.78) \[ P\{e_{n_i} > \delta/2\} \rightarrow P\{M\{\{e_j : j \in I_i\}\} > \delta/2\}, \]

and

(3.79) \[ M\{\{e_j : j \in I_i\}\} \xrightarrow{P} 0. \]

From (3.76)-(3.79), \[ P|m_i = 1|b_i < 0 \rightarrow 0. \]

Therefore, using the same argument of Theorem 3.1, we get \( \hat{\beta} \rightarrow \beta \) in probability.

QED.
4. The Asymptotic Distribution of the Estimator

In this section I study the asymptotic distribution of the estimator. The asymptotic normality of the estimator when \( x_j^0 \) take finitely many values is proved in Theorem 4.1. The proof is then extended to the general case in Theorem 4.3.

[Theorem 4.1]

Let \( x_j^0 \) be discrete and take finite different values. If:

i) \( x_j^0 \) are i.i.d.

ii) \( u_j \) are i.i.d. with median 0. The density function \( f(u) \) satisfies the following conditions.

a) \( f(u) \) is continuous and differentiable in the neighborhood of zero.

b) \( f(0) > 0 \).

c) \( f'(u) < \infty \) in the neighborhood of zero.

iii) \( \{u_j\} \) and \( \{x_j^0\} \) are independent.

iv) \( A \) is positive definite.

v) \( X_i^\beta \neq 0 \) for all \( i \).

Then

\[
\sqrt{N} (\hat{\beta} - \beta) \xrightarrow{d} N(0, \frac{1}{4f'(0)^2} A^{-1})
\]

where

\[
A = \text{plim} \sum_{N \to \infty} \left( \frac{n_i}{N} \right) X_i^1 X_i^1' = \sum_{N \to \infty} \sum_{i \in \{X_i^1: X_i^1 \beta > 0\}} p_i X_i^1 X_i^1', \quad \text{and}
\]

\[
p_i = \text{plim} \left( \frac{n_i}{N} \right)
\]
[Proof]

As I proved in Theorem 3.1,

\[ (4.2) \quad \frac{\hat{\sigma}}{\sqrt{n_1}} \to 0, \]

\[ (4.3) \quad \Pi \Phi(-\sqrt{n_1} x_1 \hat{\beta} / \hat{\sigma}) \to 1, \]

and

\[ (4.4) \quad \hat{\beta} \to \beta. \]

Let

\[ \theta = \begin{bmatrix} \alpha \\ \sigma^2 \end{bmatrix}, \quad \theta^0 = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}, \quad \hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix}. \]

Then

\[ (4.5) \quad \frac{\partial \ln L}{\partial \theta} |_{\theta^0} = \frac{\partial \ln L}{\partial \theta} |_{\hat{\theta}} + \frac{\partial^2 \ln L}{\partial \theta \partial \theta^t} |_{\theta^*}(\theta^0 - \hat{\theta}) \]

\[ = \frac{\partial^2 \ln L}{\partial \theta \partial \theta^t} |_{\theta^*}(\theta^0 - \hat{\theta}) \]

where \( \theta^* \) is some \( \Delta \) in between \( \hat{\theta} \) and \( \theta^0 \).

Here,

\[ (4.6) \quad \frac{\partial \ln L}{\partial \theta} = \begin{bmatrix} \frac{\partial \ln L}{\partial \alpha} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix}, \]
\[ (4.7) \quad \frac{\partial^2 \ln L}{\partial \theta \partial \theta'} = \begin{bmatrix} \frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma^2} \\ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \alpha'} & \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \sigma^2} \end{bmatrix}. \]

From (4.5)-(4.7),

\[ (4.8) \quad \frac{\partial \ln L}{\partial \alpha} |_{\theta^0} = \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma^2} |_{\theta^0} (\beta - \hat{\beta}) . \]

Therefore,

\[ (4.9) \quad \sqrt{N} (\hat{\beta} - \beta) = -\left[ \frac{\sigma^2}{N} \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma^2} |_{\theta^0} \right]^{-1} \left[ \frac{\sigma^2}{\sqrt{N}} \frac{\partial \ln L}{\partial \alpha} |_{\theta^0} \right] . \]

Here,

\[ (4.10) \quad \frac{\partial \ln L}{\partial \alpha} = -\sum_{i} \frac{n_i (Y_i - X'_i \alpha) X_i}{\sigma^2} - \sum_{i} \frac{\sqrt{n_i} \phi(-\sqrt{n_i} X'_i \alpha / \sigma) X_i}{\sigma \Phi(-\sqrt{n_i} X'_i \alpha / \sigma)} , \]

and

\[ (4.11) \quad \frac{\partial^2 \ln L}{\partial \alpha \partial \sigma^2} = -\sum_{i} \frac{n_i X_i X'_i}{\sigma^2} \]

\[ - \sum_{i} \frac{n_i [-\phi'(-\sqrt{n_i} X'_i \alpha / \sigma) \Phi(-\sqrt{n_i} X'_i \alpha / \sigma) + \phi^2(-\sqrt{n_i} X'_i \alpha / \sigma)]}{\sigma^2 \phi^2(-\sqrt{n_i} X'_i \alpha / \sigma)} X_i X'_i , \]

where
\[ \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} . \]

Therefore,
\[ (4.12) \quad \frac{\hat{\sigma}^2}{\sqrt{N}} \frac{\partial \ln L}{\partial \alpha} \bigg|_{\theta^{*}} = \sum_{i} n_i X_i \tilde{\epsilon}_i / \sqrt{N} \]

\[ - \frac{1}{\sqrt{N}} \sum_{i} \frac{\sqrt{n_i \Phi(-\sqrt{n_i} X_i' \beta / \sigma) X_i}}{\sigma \Phi(-\sqrt{n_i} X_i' \beta / \sigma)} \]

where \( \tilde{\epsilon}_i = M\{u_j : j \in I_i\} \),

and

\[ (4.13) \quad \frac{\hat{\sigma}^2}{N} \frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} \bigg|_{\theta^{*}} = -\sum_{i} X_i X_i' \]

\[ - \frac{1}{\sqrt{N}} \sum_{i} \frac{n_i \Phi'(-\sqrt{n_i} X_i' \beta / \sigma) \Phi(-\sqrt{n_i} X_i' \beta / \sigma) + \Phi^2(-\sqrt{n_i} X_i' \beta / \sigma)}{N \Phi^2(-\sqrt{n_i} X_i' \beta / \sigma)} X_i X_i' \]

As I show in Appendix B, the second term of (4.12) goes to zero in probability. In the same way, the second term of (4.13) also goes to zero in probability.

From Proposition 3.1,

\[ (4.14) \quad P[\sqrt{n_i} \tilde{\epsilon}_i \neq \sqrt{n_i} \epsilon_i | X_i' \beta > 0] \to 0 \]

where \( \epsilon_i = M\{u_j : j \in I_i\} \).

Since \( \sqrt{n_i} \epsilon_i = \sqrt{n_i} M\{u_j : j \in I_i\} \to N(0, 1/4f(0)^2) \),

\[ (4.15) \quad \frac{\sigma^2}{\sqrt{N}} \frac{\partial \ln L}{\partial \alpha} \bigg|_{\theta^{*}} \sim \sum_{i} n_i X_i \epsilon_i / \sqrt{N} \to N(0, \frac{1}{4f(0)^2} A) \]

and

\[ (4.16) \quad \frac{\sigma^2}{N} \frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} \bigg|_{\theta^{*}} \sim -\sum_{i} \frac{n_i}{N} X_i X_i' \to -A \]

From (4.9), (4.15) and (4.16),

\[ \sqrt{N} (\hat{\beta} - \beta) \to N(0, \frac{1}{4f(0)^2} A^{-1}). \] QED.
Next I prove the asymptotic normality of the estimator in the general case. To prove the asymptotic normality, the following theorem is used.

[Theorem 4.2]

In addition to the assumptions in Theorem 3.2, if:

i) $x_j^0$ are i.i.d. and bounded.

ii) $u_j$ are i.i.d. with median zero. The density function $f(u)$ satisfies the following conditions.
   a) $f(u)$ is continuous and differentiable in the neighborhood of zero.
   b) $f(0)>0$.
   c) $f'(u) < \infty$ in the neighborhood of zero.

iii) $\{u_j\}$ and $\{x_j^0\}$ are independent.

iv) $\{S^*_1\}$ and $\{u_j\}$ are independent.

Then for any $\varepsilon>0$, there exist $M_\varepsilon$ and $N_0$ such that

(4.18) \[ P\{\rho(\sqrt{N}(\hat{\beta}_1-\beta_1),0)> M_\varepsilon \} < \varepsilon \text{ for } N>N_0 \.

[Proof]

As shown in Theorem 3.2, $P\{M[\{u_j:j\in I_1^*\}] \# M[\{\tilde{u}_j:j\in I_1^*\}] \to 0$ as $N \to \infty$. Hence, by the same argument of Theorem 4.3 of Nawata [1985], we obtain (4.18).

QED.
Next I prove the asymptotic distribution of the estimator in the general case.

[Theorem 4.3]

In addition to the assumptions of Theorem 4.2, if $S_i$ are fixed cells and $\mathbf{A}$ is a positive definite matrix,

\begin{equation}
\sqrt{N}(\hat{\beta} - \beta) \overset{\text{d}}{\rightarrow} N(0, \frac{1}{4f(0)^2} \mathbf{A}^{-1})
\end{equation}

where

$$
\mathbf{A} = \lim_{N \to \infty} \frac{1}{N} \sum_{i} n_i X_i X_i' = \sum_{i \in \{i | b_i > 0\}} p_i \xi_i X_i' X_i
$$

$$
p_i = \lim(n_i / N) \text{ and } X_i^* = \lim X_i = \mathbf{E}[x_j | \xi I_1].
$$

[Proof]

Using the same argument as in Theorem 4.1, we get,

\begin{equation}
\sqrt{N}(\hat{\beta} - \beta) = -[\sigma^2 N \frac{\partial^2 \ln L_1}{\partial \delta \partial \alpha'}]^{-1} \sigma^2 N \frac{\partial \ln L_1}{\partial \alpha} \theta^0
\end{equation}

where $\theta^0 = \begin{bmatrix} \beta \\ \sigma^2 \end{bmatrix}$, $\hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2 \end{bmatrix}$, and $\theta^*$ is some value between $\theta^0$ and $\hat{\theta}$.

Here,

\begin{equation}
\frac{\hat{\sigma}^2}{\sqrt{N}} \frac{\partial \ln L_1}{\partial \alpha} \bigg|_{\theta^0} \overset{\text{d}}{\rightarrow} \sum_{i} n_i X_i \tilde{e}_i / \sqrt{N},
\end{equation}

and

\begin{equation}
\frac{\hat{\sigma}^2}{N} \frac{\partial^2 \ln L_1}{\partial \delta \partial \alpha'} \bigg|_{\theta^*} \overset{\text{d}}{\rightarrow} \sum_{i} n_i X_i X_i'
\end{equation}

where $\tilde{e}_i = N\{(x_i^0 - x_j^0)'(\beta_i^0 - \beta_j^0) + u_j : j \in I_1\} = M\{e_j : j \in I_1\}$, and
\[ \tilde{\varepsilon}_i = M[\{(x_i^0-x_j^0)'(\bar{\beta}_i-\beta_i)+u_i;j\varepsilon I_1\}] = M[\{(\tilde{\varepsilon}_i;j\varepsilon I_1\}] \]

From Proposition 3.1, for \( \beta_i>0 \),

\[
(4.23) \quad \sqrt{n}_i (\tilde{\varepsilon}_i - \varepsilon_i) \xrightarrow{p} 0 .
\]

From Theorem 4.2 of Nawata [1985],

\[
(4.24) \quad \sqrt{n}_i \varepsilon_i \xrightarrow{d} N(0, \frac{1}{4f(0)^2}) .
\]

Since the \( \varepsilon_i \)'s are independent,

\[
(4.25) \quad \sqrt{n}_i (\hat{\beta}-\beta) \xrightarrow{d} N(0, \frac{1}{4f(0)^2} A^{-1}) .
\]

QED.

Theorems 4.1 and 4.3 suggest the following:

i) The asymptotic variance of the estimator is determined by the cells in which \( x_j^0 \beta > 0 \) for all \( j \). The cells which contain \( x_j^0 \beta < 0 \) do not affect the asymptotic distribution of the estimator.

ii) When \( x_j^0 \) are discrete and take a finite number of different values, this estimator and Powell's LAD estimator have the same asymptotic distributions.

iii) When \( x_j^0 \) take an infinite number of different values, the asymptotic variance of this estimator approaches Powell's as we make the size of each cell smaller. So we can make
the asymptotic variance of the estimator arbitrarily close to Powell's. In this sense, this estimator is asymptotically as efficient as Powell's LAD estimator.
Monte Carlo Study

The results of the Monte Carlo study are presented in this section. The effects of nonnormality and heteroscedasticity are studied. This study is styled after the study of Paarsch [1983].

The basic model of this study is:

\[ y_j = a + b x_j + \tilde{u}_j \]  

where \[ \tilde{u}_j = \begin{cases} 
  u_j & \text{if } u_j > -(a + b x_j) \\
  -(a + b x_j) & \text{otherwise} 
\end{cases} \]

5.1 Nonnormality

In this section, the following items are studied:

i) The effect of different distributions. The Cauchy, Laplace, and normal distributions are considered.

ii) The effect of sample size. Sample sizes of fifty, one hundred, and two hundred are considered.

iii) The effect of the degree of censoring. Twenty-five percent and fifty percent are considered.

The values of \( x_j^0 \) are chosen from \([0, 20]\), spacing them in equidistant points from 0 to 20, where the distance between the points is determined by the sample size. The value of \( b \) is 1.0 for all cases. For the 50% censored cases, \( a = 10.0 \). For the 25% censored cases, \( a = -1.60 \) for the Cauchy distribution; \( a = -2.09 \) for the Laplace distribution; and \( a = -2.94 \) for the normal distribution. For the Laplace and normal cases, the expected value of the error term is 0 and its variance is
100. For the Cauchy cases, the location parameter is 0 and the scale parameter is 10.

The Tobit MLE, Powell's LAD, $\hat{\beta}$, and $\hat{\beta}^*$ which maximizes

$$L(\alpha, \sigma) = \prod_{m_1=1}^{\sqrt{n_1} \sigma^{-1}} \exp{-\frac{(Y_i - X_i'\alpha)^2}{2\sigma^2}}$$

are studied. $\hat{\beta}^*$ is the WLS estimator based on $m_1=1$ cells. The numbers of cells are 16 for the 50-sample cases and 20 for the 100- and 200-sample cases. The size of cell is chosen to be equal and is determined by the number of cells. The number of trials is 100 per case. The Tobit MLE and Powell's LAD are calculated by Paarsch [1983].

The findings of the study are:

i) Although it is the best estimator when the distribution of the error term is Laplace or normal, the Tobit MLE is the poorest estimator for the Cauchy distribution.

ii) For the 50% censored cases, $\hat{\beta}$ is much better than Powell's LAD estimator for all the distributions. For the 25% censored cases, the performances of these two estimators are similar.

iii) Although $\hat{\beta}$ and $\hat{\beta}^*$ have the same asymptotic properties, the performance of $\hat{\beta}$ is much better than $\hat{\beta}^*$.

iv) $\hat{\beta}^*$ overestimates $\alpha$ and underestimates $\beta$. 
### TABLE 1
CAUCHY
SAMPLE SIZE=50
FIFTY PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-28.74</td>
<td>592.21</td>
<td>-94.72</td>
<td>-35.92</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>4.90</td>
<td>61.91</td>
<td>1.28</td>
<td>2.34</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-17.68</td>
<td>17.13</td>
<td>-28.51</td>
<td>-11.18</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.25</td>
<td>1.44</td>
<td>0.66</td>
<td>1.20</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>a</td>
<td>-10.00</td>
<td>-15.49</td>
<td>17.00</td>
<td>-19.07</td>
<td>-11.48</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.35</td>
<td>1.09</td>
<td>0.78</td>
<td>1.23</td>
</tr>
<tr>
<td>( \hat{\beta}^* )</td>
<td>a</td>
<td>-10.00</td>
<td>10.51</td>
<td>28.21</td>
<td>-0.06</td>
<td>6.22</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.11</td>
<td>1.77</td>
<td>-0.22</td>
<td>0.38</td>
</tr>
</tbody>
</table>

### TABLE 2
LAPLACE
SAMPLE SIZE=50
FIFTY PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-11.74</td>
<td>4.68</td>
<td>-14.31</td>
<td>-11.66</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.13</td>
<td>0.35</td>
<td>0.86</td>
<td>1.09</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-14.59</td>
<td>11.90</td>
<td>-18.52</td>
<td>-10.53</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.27</td>
<td>0.69</td>
<td>0.69</td>
<td>1.11</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>a</td>
<td>-10.00</td>
<td>-11.78</td>
<td>5.67</td>
<td>-15.04</td>
<td>-11.39</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.12</td>
<td>0.41</td>
<td>0.82</td>
<td>1.11</td>
</tr>
<tr>
<td>( \hat{\beta}^* )</td>
<td>a</td>
<td>-10.00</td>
<td>1.39</td>
<td>7.21</td>
<td>-2.54</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.45</td>
<td>0.48</td>
<td>0.19</td>
<td>0.45</td>
</tr>
</tbody>
</table>
### TABLE 3
NORMAL
SAMPLE SIZE=50
FIFTY PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-9.92</td>
<td>4.26</td>
<td>-12.41</td>
<td>-10.12</td>
<td>-6.52</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.00</td>
<td>0.31</td>
<td>0.76</td>
<td>1.03</td>
<td>1.22</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-15.55</td>
<td>13.83</td>
<td>-20.82</td>
<td>-11.03</td>
<td>-5.30</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.34</td>
<td>0.81</td>
<td>0.78</td>
<td>1.16</td>
<td>1.66</td>
</tr>
<tr>
<td>ß</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-10.88</td>
<td>5.62</td>
<td>-14.11</td>
<td>-10.76</td>
<td>-6.85</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.06</td>
<td>0.39</td>
<td>0.79</td>
<td>1.07</td>
<td>1.28</td>
</tr>
<tr>
<td>ß*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>0.78</td>
<td>5.67</td>
<td>-2.32</td>
<td>0.32</td>
<td>3.84</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.48</td>
<td>0.40</td>
<td>0.26</td>
<td>0.51</td>
<td>0.72</td>
</tr>
</tbody>
</table>

### TABLE 4
CAUCHY
SAMPLE SIZE=100
FIFTY PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-108.33</td>
<td>423.46</td>
<td>-121.30</td>
<td>-57.66</td>
<td>-31.61</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>7.30</td>
<td>37.21</td>
<td>1.94</td>
<td>3.90</td>
<td>7.55</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-16.73</td>
<td>13.52</td>
<td>-22.29</td>
<td>-12.83</td>
<td>-6.24</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.42</td>
<td>0.86</td>
<td>0.78</td>
<td>1.23</td>
<td>1.86</td>
</tr>
<tr>
<td>ß</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-10.70</td>
<td>6.61</td>
<td>-13.55</td>
<td>-8.74</td>
<td>-6.20</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.54</td>
<td>0.67</td>
<td>0.95</td>
<td>1.28</td>
</tr>
<tr>
<td>ß*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>3.87</td>
<td>8.66</td>
<td>-0.43</td>
<td>3.12</td>
<td>7.90</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.34</td>
<td>0.61</td>
<td>-0.03</td>
<td>0.42</td>
<td>0.71</td>
</tr>
<tr>
<td>TABLE 5</td>
<td>LAPLACE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAMPLE SIZE=100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIFTY PERCENT CENSORING</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>truth</td>
<td>mean</td>
<td>s.d.</td>
<td>25%</td>
<td>median</td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
<td>-------</td>
<td>-----</td>
<td>--------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Tobit a</td>
<td>-10.00</td>
<td>-12.09</td>
<td>3.59</td>
<td>-14.35</td>
<td>-11.70</td>
<td>-9.68</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.14</td>
<td>0.27</td>
<td>0.96</td>
<td>1.10</td>
<td>1.33</td>
</tr>
<tr>
<td>Powell a</td>
<td>-10.00</td>
<td>-12.78</td>
<td>7.02</td>
<td>-15.95</td>
<td>-10.99</td>
<td>-7.75</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.17</td>
<td>0.44</td>
<td>0.86</td>
<td>1.12</td>
<td>1.33</td>
</tr>
<tr>
<td>( \hat{\beta} ) a</td>
<td>-10.00</td>
<td>-9.87</td>
<td>4.05</td>
<td>-12.24</td>
<td>-9.22</td>
<td>-6.78</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.99</td>
<td>0.31</td>
<td>0.75</td>
<td>0.95</td>
<td>1.17</td>
</tr>
<tr>
<td>( \hat{\beta}^* ) a</td>
<td>-10.00</td>
<td>-0.67</td>
<td>4.77</td>
<td>-2.88</td>
<td>-0.03</td>
<td>2.22</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.50</td>
<td>0.35</td>
<td>0.27</td>
<td>0.49</td>
<td>0.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 6</th>
<th>NORMAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMPLE SIZE=100</td>
<td></td>
</tr>
<tr>
<td>FIFTY PERCENT CENSORING</td>
<td></td>
</tr>
<tr>
<td>truth</td>
<td>mean</td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Tobit a</td>
<td>-10.00</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
</tr>
<tr>
<td>Powell a</td>
<td>-10.00</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\beta} ) a</td>
<td>-10.00</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
</tr>
<tr>
<td>( \hat{\beta}^* ) a</td>
<td>-10.00</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
</tr>
</tbody>
</table>
### TABLE 7
**CAUCHY**
**SAMPLE SIZE=200**
**FIFTY PERCENT CENSORING**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>143.02</td>
<td>2419.13</td>
<td>-152.03</td>
<td>-74.77</td>
<td>-48.84</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>-16.38</td>
<td>190.38</td>
<td>2.46</td>
<td>1.11</td>
<td>1.63</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-14.10</td>
<td>9.92</td>
<td>-19.91</td>
<td>-11.32</td>
<td>-6.47</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.25</td>
<td>0.61</td>
<td>0.76</td>
<td>1.11</td>
<td>1.63</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-11.12</td>
<td>5.45</td>
<td>-13.72</td>
<td>-9.85</td>
<td>-7.45</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.07</td>
<td>0.38</td>
<td>0.82</td>
<td>1.02</td>
<td>1.27</td>
</tr>
<tr>
<td>( \hat{\beta}^* )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>0.09</td>
<td>5.17</td>
<td>-3.22</td>
<td>-0.17</td>
<td>2.38</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.50</td>
<td>0.34</td>
<td>0.29</td>
<td>0.51</td>
<td>0.71</td>
</tr>
</tbody>
</table>

### TABLE 8
**LAPLACE**
**SAMPLE SIZE=200**
**FIFTY PERCENT CENSORING**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-11.94</td>
<td>2.26</td>
<td>-13.29</td>
<td>-11.90</td>
<td>-10.28</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.13</td>
<td>0.17</td>
<td>1.01</td>
<td>1.12</td>
<td>1.25</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-11.07</td>
<td>4.38</td>
<td>-13.10</td>
<td>-10.30</td>
<td>-8.14</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.07</td>
<td>0.28</td>
<td>0.90</td>
<td>1.02</td>
<td>1.21</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-10.26</td>
<td>3.58</td>
<td>-12.23</td>
<td>-9.21</td>
<td>-7.93</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.25</td>
<td>0.86</td>
<td>0.96</td>
<td>1.17</td>
</tr>
<tr>
<td>( \hat{\beta}^* )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-3.27</td>
<td>4.18</td>
<td>-5.59</td>
<td>-2.83</td>
<td>-0.93</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.64</td>
<td>0.27</td>
<td>0.49</td>
<td>0.62</td>
<td>0.81</td>
</tr>
</tbody>
</table>
### Table 9
**Normal**
**Sample Size = 200**
**Fifty Percent Censoring**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tobit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-10.26</td>
<td>2.03</td>
<td>-11.87</td>
<td>-10.32</td>
<td>-8.95</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.15</td>
<td>0.91</td>
<td>1.00</td>
<td>1.11</td>
</tr>
<tr>
<td><strong>Powell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-12.48</td>
<td>7.58</td>
<td>-15.55</td>
<td>-10.72</td>
<td>-7.04</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.15</td>
<td>0.47</td>
<td>0.78</td>
<td>1.00</td>
<td>1.37</td>
</tr>
<tr>
<td><strong>β</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-10.33</td>
<td>3.76</td>
<td>-12.14</td>
<td>-9.36</td>
<td>-7.84</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.26</td>
<td>0.84</td>
<td>0.97</td>
<td>1.18</td>
</tr>
<tr>
<td><strong>β</strong>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.00</td>
<td>-2.93</td>
<td>4.00</td>
<td>-5.06</td>
<td>-2.52</td>
<td>-0.69</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.62</td>
<td>0.26</td>
<td>0.47</td>
<td>0.63</td>
<td>0.77</td>
</tr>
</tbody>
</table>
TABLE 10
CAUCHY
SAMPLE SIZE=50
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>1.60</td>
<td>-329.02</td>
<td>3668.57</td>
<td>-16.19</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>25.73</td>
<td>300.96</td>
<td>0.53</td>
<td>1.22</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>1.60</td>
<td>0.11</td>
<td>8.81</td>
<td>-2.41</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.13</td>
<td>0.63</td>
<td>0.73</td>
<td>1.03</td>
</tr>
<tr>
<td>β</td>
<td>a</td>
<td>1.60</td>
<td>2.08</td>
<td>7.18</td>
<td>-2.08</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.99</td>
<td>0.65</td>
<td>0.60</td>
<td>0.94</td>
</tr>
<tr>
<td>β*</td>
<td>a</td>
<td>1.60</td>
<td>8.71</td>
<td>9.73</td>
<td>4.07</td>
<td>7.79</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.74</td>
<td>0.73</td>
<td>0.45</td>
<td>0.82</td>
</tr>
</tbody>
</table>

TABLE 11
LAPLACE
SAMPLE SIZE=50
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-2.94</td>
<td>-3.24</td>
<td>2.96</td>
<td>-5.66</td>
<td>-2.89</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.04</td>
<td>0.25</td>
<td>0.86</td>
<td>1.01</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-2.94</td>
<td>-3.64</td>
<td>4.53</td>
<td>-4.67</td>
<td>-2.85</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.05</td>
<td>0.31</td>
<td>0.87</td>
<td>1.01</td>
</tr>
<tr>
<td>β</td>
<td>a</td>
<td>-2.94</td>
<td>-2.72</td>
<td>4.05</td>
<td>-5.57</td>
<td>-2.41</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.00</td>
<td>0.36</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>β*</td>
<td>a</td>
<td>-2.94</td>
<td>3.04</td>
<td>4.12</td>
<td>0.50</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.72</td>
<td>0.32</td>
<td>0.51</td>
<td>0.73</td>
</tr>
</tbody>
</table>
### TABLE 12
NORMAL
SAMPLE SIZE=50
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-2.12</td>
<td>3.01</td>
<td>-4.49</td>
<td>-1.82</td>
<td>-0.29</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.01</td>
<td>0.24</td>
<td>0.83</td>
<td>0.99</td>
<td>1.16</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-1.43</td>
<td>7.94</td>
<td>-4.66</td>
<td>-1.99</td>
<td>0.12</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.12</td>
<td>0.52</td>
<td>0.78</td>
<td>1.02</td>
<td>1.25</td>
</tr>
<tr>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-2.09</td>
<td>4.10</td>
<td>-5.03</td>
<td>-1.77</td>
<td>0.62</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.01</td>
<td>0.33</td>
<td>0.75</td>
<td>1.02</td>
<td>1.28</td>
</tr>
<tr>
<td>β*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>2.58</td>
<td>3.59</td>
<td>0.26</td>
<td>2.82</td>
<td>4.60</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.69</td>
<td>0.28</td>
<td>0.53</td>
<td>0.69</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### TABLE 13
CAUCHY
SAMPLE SIZE=100
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1.60</td>
<td>-226.76</td>
<td>2807.60</td>
<td>-25.41</td>
<td>-7.49</td>
<td>6.34</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>22.28</td>
<td>222.67</td>
<td>0.74</td>
<td>1.75</td>
<td>3.03</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1.60</td>
<td>0.98</td>
<td>4.19</td>
<td>-0.90</td>
<td>1.33</td>
<td>3.31</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.06</td>
<td>0.33</td>
<td>0.86</td>
<td>1.05</td>
<td>1.19</td>
</tr>
<tr>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1.60</td>
<td>2.47</td>
<td>3.30</td>
<td>-0.06</td>
<td>2.31</td>
<td>4.73</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.05</td>
<td>0.30</td>
<td>0.75</td>
<td>0.94</td>
<td>1.13</td>
</tr>
<tr>
<td>β*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1.60</td>
<td>5.95</td>
<td>4.42</td>
<td>2.79</td>
<td>6.02</td>
<td>9.03</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.78</td>
<td>0.37</td>
<td>0.54</td>
<td>0.73</td>
<td>1.06</td>
</tr>
</tbody>
</table>
### TABLE 14
LAPLACE
SAMPLE SIZE=100
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-2.94</td>
<td>3.22</td>
<td>2.10</td>
<td>-4.38</td>
<td>-3.36</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.04</td>
<td>0.18</td>
<td>0.92</td>
<td>1.02</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-2.94</td>
<td>3.44</td>
<td>2.25</td>
<td>-4.57</td>
<td>-2.90</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.04</td>
<td>0.17</td>
<td>0.92</td>
<td>1.03</td>
</tr>
<tr>
<td>β</td>
<td>a</td>
<td>-2.94</td>
<td>2.06</td>
<td>2.29</td>
<td>-3.53</td>
<td>-2.10</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.94</td>
<td>0.20</td>
<td>0.82</td>
<td>0.93</td>
</tr>
<tr>
<td>β*</td>
<td>a</td>
<td>-2.94</td>
<td>1.55</td>
<td>2.90</td>
<td>-0.86</td>
<td>1.79</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.78</td>
<td>0.26</td>
<td>0.60</td>
<td>0.77</td>
</tr>
</tbody>
</table>

### TABLE 15
NORMAL
SAMPLE SIZE=100
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-2.09</td>
<td>-2.05</td>
<td>2.06</td>
<td>-3.43</td>
<td>-2.13</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.00</td>
<td>0.18</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-2.09</td>
<td>-2.91</td>
<td>3.73</td>
<td>-4.83</td>
<td>-1.94</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.07</td>
<td>0.28</td>
<td>0.87</td>
<td>1.06</td>
</tr>
<tr>
<td>β</td>
<td>a</td>
<td>-2.09</td>
<td>-1.31</td>
<td>2.38</td>
<td>-2.80</td>
<td>-1.28</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.95</td>
<td>0.21</td>
<td>0.82</td>
<td>0.94</td>
</tr>
<tr>
<td>β*</td>
<td>a</td>
<td>-2.09</td>
<td>1.30</td>
<td>2.81</td>
<td>-0.41</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.74</td>
<td>0.25</td>
<td>0.59</td>
<td>0.71</td>
</tr>
</tbody>
</table>
### TABLE 16
CAUCHY
SAMPLE SIZE=200
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>1.60</td>
<td>-279.81</td>
<td>2302.10</td>
<td>-32.88</td>
<td>-12.98</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>-19.42</td>
<td>163.85</td>
<td>0.88</td>
<td>1.75</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>1.60</td>
<td>1.33</td>
<td>2.87</td>
<td>-0.11</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.22</td>
<td>0.86</td>
<td>0.97</td>
</tr>
<tr>
<td>(\hat{\beta})</td>
<td>a</td>
<td>1.60</td>
<td>1.71</td>
<td>1.99</td>
<td>0.52</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.99</td>
<td>0.18</td>
<td>0.87</td>
<td>0.98</td>
</tr>
<tr>
<td>(\hat{\beta}^*)</td>
<td>a</td>
<td>1.60</td>
<td>3.78</td>
<td>2.82</td>
<td>1.97</td>
<td>3.76</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.88</td>
<td>0.24</td>
<td>0.68</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### TABLE 17
LAPLACE
SAMPLE SIZE=200
TWENTY-FIVE PERCENT CENSORING

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-2.94</td>
<td>-3.26</td>
<td>1.56</td>
<td>-4.52</td>
<td>-2.98</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.04</td>
<td>0.13</td>
<td>0.94</td>
<td>1.03</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-2.94</td>
<td>-3.17</td>
<td>1.56</td>
<td>-3.84</td>
<td>-2.90</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.02</td>
<td>0.12</td>
<td>0.94</td>
<td>1.00</td>
</tr>
<tr>
<td>(\hat{\beta})</td>
<td>a</td>
<td>-2.94</td>
<td>-2.66</td>
<td>1.56</td>
<td>-2.79</td>
<td>-1.87</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.97</td>
<td>0.14</td>
<td>0.89</td>
<td>0.99</td>
</tr>
<tr>
<td>(\hat{\beta}^*)</td>
<td>a</td>
<td>-2.94</td>
<td>0.09</td>
<td>1.96</td>
<td>-1.13</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.87</td>
<td>0.16</td>
<td>0.76</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>truth</td>
<td>mean</td>
<td>s.d.</td>
<td>25%</td>
<td>median</td>
<td>75%</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Tobit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-2.15</td>
<td>1.66</td>
<td>-3.60</td>
<td>-1.95</td>
<td>-1.01</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.01</td>
<td>0.13</td>
<td>0.91</td>
<td>1.00</td>
<td>1.11</td>
</tr>
<tr>
<td>Powell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-2.59</td>
<td>2.54</td>
<td>-3.52</td>
<td>-2.08</td>
<td>-1.05</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.04</td>
<td>0.19</td>
<td>0.91</td>
<td>1.00</td>
<td>1.15</td>
</tr>
<tr>
<td>β</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-1.89</td>
<td>1.56</td>
<td>-2.79</td>
<td>-1.87</td>
<td>-0.94</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.99</td>
<td>0.14</td>
<td>0.89</td>
<td>0.99</td>
<td>1.09</td>
</tr>
<tr>
<td>β*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.09</td>
<td>-0.18</td>
<td>2.00</td>
<td>-1.44</td>
<td>-0.03</td>
<td>1.12</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.83</td>
<td>0.16</td>
<td>0.73</td>
<td>0.83</td>
<td>0.91</td>
</tr>
</tbody>
</table>

s.d.: standard error of the estimator
25%: 25% quantile
75%: 75% quantile
5.2 Heteroscedasticity

The effect of heteroscedasticity is studied in this section. The values of $x_j^2$ are chosen in the same way as 5.1. The error terms are normally distributed with mean zero. The variance is given by:

\[(5.3) \quad \text{Var } u_j = 100*(1+x_j^2/20)^2,\]

or

\[(5.4) \quad \text{var } u_j = 100*(2-x_j^2/20)^2.\]

The value of $b$ is 1.0 for all the cases and $a$ is either -10.0 or -2.94 as before.

The findings of the study are:

i) If the variance is (5.3), Tobit underestimates $a$ and overestimates $b$. If the variance is (5.4), Tobit overestimates $a$ and underestimates $b$.

ii) The Tobit estimator has a fairly large bias, although it has the smallest variance.

iii) The bias of the Tobit estimator does not become small as the number of observations increases.

iv) The bias of $\hat{\beta}$ is smaller than Tobit in all cases, although the variance of $\hat{\beta}$ is usually a little bigger than Tobit.

v) The bias of $\hat{\beta}$ and Powell's estimator becomes smaller as the number of observations increases.

vi) Powell's estimator performs poorly, especially when the degree of censoring is high (when $a=-10.0$).
TABLE 19
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=50

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-13.34</td>
<td>6.91</td>
<td>-18.30</td>
<td>-13.89</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.31</td>
<td>0.54</td>
<td>0.92</td>
<td>1.31</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-25.50</td>
<td>19.06</td>
<td>-47.77</td>
<td>-21.00</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.92</td>
<td>1.12</td>
<td>0.99</td>
<td>1.71</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>a</td>
<td>-10.00</td>
<td>-13.77</td>
<td>9.70</td>
<td>-10.79</td>
<td>-12.09</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.20</td>
<td>0.63</td>
<td>0.79</td>
<td>1.16</td>
</tr>
</tbody>
</table>

TABLE 20
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=100

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-14.43</td>
<td>4.97</td>
<td>-17.57</td>
<td>-13.74</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.31</td>
<td>0.40</td>
<td>1.02</td>
<td>1.29</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-10.71</td>
<td>17.25</td>
<td>-32.27</td>
<td>-12.41</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.50</td>
<td>1.02</td>
<td>0.65</td>
<td>1.28</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>a</td>
<td>-10.00</td>
<td>-12.71</td>
<td>6.47</td>
<td>-16.63</td>
<td>-12.58</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.18</td>
<td>0.51</td>
<td>0.85</td>
<td>1.16</td>
</tr>
</tbody>
</table>

TABLE 21
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=200

<table>
<thead>
<tr>
<th></th>
<th>true</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-14.91</td>
<td>3.30</td>
<td>-16.64</td>
<td>-14.53</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.36</td>
<td>0.29</td>
<td>1.15</td>
<td>1.32</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-10.96</td>
<td>16.31</td>
<td>-25.77</td>
<td>-11.98</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.54</td>
<td>0.94</td>
<td>0.82</td>
<td>1.26</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>a</td>
<td>-10.00</td>
<td>-12.96</td>
<td>5.95</td>
<td>-15.34</td>
<td>-11.64</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.21</td>
<td>0.43</td>
<td>0.92</td>
<td>1.18</td>
</tr>
</tbody>
</table>
TABLE 22
Var \( u_i = 100(2-x_i^2)/20 \), SAMPLE SIZE=50

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-6.32</td>
<td>5.76</td>
<td>-10.13</td>
<td>-5.77</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.74</td>
<td>0.42</td>
<td>0.43</td>
<td>0.74</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-19.20</td>
<td>16.50</td>
<td>-27.96</td>
<td>-13.65</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.52</td>
<td>0.99</td>
<td>0.92</td>
<td>1.27</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>a</td>
<td>-10.00</td>
<td>-7.89</td>
<td>5.97</td>
<td>-10.55</td>
<td>-7.15</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.87</td>
<td>0.42</td>
<td>0.57</td>
<td>0.83</td>
</tr>
</tbody>
</table>

TABLE 23
Var \( u_i = 100(2-x_i^2)/20 \), SAMPLE SIZE=100

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-6.41</td>
<td>3.94</td>
<td>-8.92</td>
<td>-5.62</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.74</td>
<td>0.31</td>
<td>0.52</td>
<td>0.75</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-15.58</td>
<td>14.43</td>
<td>-21.50</td>
<td>-10.59</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.22</td>
<td>0.62</td>
<td>0.77</td>
<td>1.08</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>a</td>
<td>-10.00</td>
<td>-7.54</td>
<td>5.43</td>
<td>-10.59</td>
<td>-6.73</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.84</td>
<td>0.39</td>
<td>0.53</td>
<td>0.84</td>
</tr>
</tbody>
</table>

TABLE 24
Var \( u_i = 100(2-x_i^2)/20 \), SAMPLE SIZE=200

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit</td>
<td>a</td>
<td>-10.00</td>
<td>-6.64</td>
<td>2.53</td>
<td>-8.02</td>
<td>-6.41</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.76</td>
<td>0.21</td>
<td>0.61</td>
<td>0.74</td>
</tr>
<tr>
<td>Powell</td>
<td>a</td>
<td>-10.00</td>
<td>-13.41</td>
<td>10.31</td>
<td>-17.11</td>
<td>-9.84</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.22</td>
<td>0.61</td>
<td>0.77</td>
<td>1.08</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>a</td>
<td>-10.00</td>
<td>-8.12</td>
<td>4.02</td>
<td>-9.32</td>
<td>-7.61</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>0.90</td>
<td>0.28</td>
<td>0.74</td>
<td>0.87</td>
</tr>
</tbody>
</table>
TABLE 25  
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=50  

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit a</td>
<td>-2.94</td>
<td>-5.02</td>
<td>5.62</td>
<td>-9.22</td>
<td>-5.13</td>
<td>-0.21</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.16</td>
<td>0.48</td>
<td>0.82</td>
<td>1.17</td>
</tr>
<tr>
<td>Powell a</td>
<td>-2.94</td>
<td>-8.73</td>
<td>12.80</td>
<td>-11.32</td>
<td>-4.04</td>
<td>-0.47</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.31</td>
<td>0.87</td>
<td>0.84</td>
<td>1.03</td>
</tr>
<tr>
<td>$\hat{\beta}$ a</td>
<td>-2.94</td>
<td>-4.53</td>
<td>7.05</td>
<td>-8.94</td>
<td>-4.09</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.09</td>
<td>0.53</td>
<td>0.78</td>
<td>1.12</td>
</tr>
</tbody>
</table>

TABLE 26  
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=100  

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit a</td>
<td>-2.94</td>
<td>-5.17</td>
<td>3.91</td>
<td>-7.53</td>
<td>-4.45</td>
<td>-2.18</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.16</td>
<td>0.36</td>
<td>0.90</td>
<td>1.14</td>
</tr>
<tr>
<td>Powell a</td>
<td>-2.94</td>
<td>-5.19</td>
<td>7.99</td>
<td>-7.11</td>
<td>-2.77</td>
<td>-0.65</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.36</td>
<td>0.59</td>
<td>0.75</td>
<td>1.03</td>
</tr>
<tr>
<td>$\hat{\beta}$ a</td>
<td>-2.94</td>
<td>-4.09</td>
<td>4.92</td>
<td>-7.69</td>
<td>-3.60</td>
<td>-0.55</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.07</td>
<td>0.43</td>
<td>0.79</td>
<td>1.01</td>
</tr>
</tbody>
</table>

TABLE 27  
Var $u_i = 100 \times (1 + x_i^2 / 20)^2$, SAMPLE SIZE=200  

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobit a</td>
<td>-2.94</td>
<td>-5.51</td>
<td>2.66</td>
<td>-7.13</td>
<td>-5.27</td>
<td>-3.57</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.20</td>
<td>0.25</td>
<td>0.98</td>
<td>1.20</td>
</tr>
<tr>
<td>Powell a</td>
<td>-2.94</td>
<td>-5.05</td>
<td>5.82</td>
<td>-5.82</td>
<td>-5.85</td>
<td>-1.20</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.17</td>
<td>0.44</td>
<td>0.88</td>
<td>1.09</td>
</tr>
<tr>
<td>$\hat{\beta}$ a</td>
<td>-2.94</td>
<td>-4.00</td>
<td>3.43</td>
<td>-5.98</td>
<td>-3.56</td>
<td>-1.69</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.00</td>
<td>1.09</td>
<td>0.30</td>
<td>0.88</td>
<td>1.03</td>
</tr>
</tbody>
</table>
**TABLE 28**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tobit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-0.53</td>
<td>5.62</td>
<td>-4.58</td>
<td>-0.97</td>
<td>4.00</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.84</td>
<td>0.41</td>
<td>0.55</td>
<td>0.85</td>
<td>1.09</td>
</tr>
<tr>
<td><strong>Powell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-6.60</td>
<td>10.94</td>
<td>-10.71</td>
<td>-4.58</td>
<td>-0.23</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.24</td>
<td>0.68</td>
<td>0.83</td>
<td>1.17</td>
<td>1.47</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-1.79</td>
<td>5.09</td>
<td>-5.21</td>
<td>-2.16</td>
<td>2.78</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.92</td>
<td>0.37</td>
<td>0.67</td>
<td>0.92</td>
<td>1.22</td>
</tr>
</tbody>
</table>

**TABLE 29**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tobit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-0.71</td>
<td>3.82</td>
<td>-3.38</td>
<td>-0.35</td>
<td>2.33</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.84</td>
<td>0.30</td>
<td>0.64</td>
<td>0.83</td>
<td>1.08</td>
</tr>
<tr>
<td><strong>Powell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-3.83</td>
<td>6.75</td>
<td>-6.43</td>
<td>-2.10</td>
<td>-0.36</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.07</td>
<td>0.45</td>
<td>0.79</td>
<td>1.01</td>
<td>1.31</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-1.09</td>
<td>4.73</td>
<td>-5.10</td>
<td>-1.60</td>
<td>1.10</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.93</td>
<td>0.34</td>
<td>0.68</td>
<td>0.90</td>
<td>1.22</td>
</tr>
</tbody>
</table>

**TABLE 30**

<table>
<thead>
<tr>
<th></th>
<th>truth</th>
<th>mean</th>
<th>s.d.</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tobit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-0.85</td>
<td>2.62</td>
<td>-2.50</td>
<td>-0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.86</td>
<td>0.21</td>
<td>0.70</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Powell</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-3.93</td>
<td>5.06</td>
<td>-5.68</td>
<td>-2.74</td>
<td>-0.91</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>1.08</td>
<td>0.35</td>
<td>0.85</td>
<td>1.01</td>
<td>1.21</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-2.94</td>
<td>-2.37</td>
<td>3.19</td>
<td>-4.53</td>
<td>-2.26</td>
<td>-0.22</td>
</tr>
<tr>
<td>b</td>
<td>1.00</td>
<td>0.98</td>
<td>0.24</td>
<td>0.82</td>
<td>0.95</td>
<td>1.14</td>
</tr>
</tbody>
</table>
s.d.: standard error of the estimator
25% : 25% quantile
75% : 75% quantile
Appendix A

The symbols used in this paper are:

1. $N$: number of observations.
2. $K$: number of independent variables (excluding the constant term).
3. $L$: number of cells at the second-stage.
4. $n_i$: number of the observations in the $i$-th cell.
5. $n_i^+$: number of positive observations in the $i$-th cell.
6. $x_j^0$: independent variables (excluding the constant).
7. $x_j^i = (1, x_j^0, i)$.
8. $y_j$: censored dependent variable ($y_j > 0$) and
   \[ y_j = \beta_0 + x_j^0 \beta_1 + u_j \]
   \[ = x_j^i \beta + u_j. \]
9. $\beta' = (\beta_0, \beta_1')$.
10. $I_i$: index set for the $i$-th cell; that is, if $j \in I_i$, $x_j^0$ belongs to the $i$-th cell.
11. $x_1^0 = \sum_{j \in I_1} x_j^0/n_i$.
12. $x_1^i = (1, x_1^0, i)$.
13. $\bar{\beta}_1$: first-stage estimator.
14. $y_j(\bar{\beta}_1)$: adjusted dependent variable of $y_j$. Let $j \in I_i$. Then,
   \[ y_j(\bar{\beta}_1) = y_j - (x_j^0 - x_1^0)' \bar{\beta}_1 \]
   \[ = x_j^i \beta + (x_j^0 - x_1^0)'(\bar{\beta}_1 - \beta_1) + u_j \]
   \[ = x_j^i \beta + e_j. \]

Let
\[y(1)(\tilde{\beta}_1) > y(2)(\tilde{\beta}_1) > \ldots > y(n_1)(\tilde{\beta}_1).\]

15. \[e_j = (x_i^0 - x_j^0)'(\tilde{\beta}_1 - \beta_1) + u_j.\]

16. \[\tilde{e}_j = (x_i^0 - x_j^0)'(\tilde{\beta}_1 - \beta_1) + \tilde{u}_j.\]

As a consequence of 14, \[\tilde{e}_1 > \tilde{e}_2 > \ldots > \tilde{e}_{n_1}.\]

17. \[M\{u_j\} = \text{sample median of the set }\{u_j\},\]

\[M\{u_j\} = \begin{cases} 
u((n+1)/2) & \text{if } n \text{ is odd} \\ [u(n/2) + u(n/2+1)]/2 & \text{if } n \text{ is even} \end{cases}\]

where \[u(1) < u(2) < \ldots < u(n).\]

18. \[e_1 = M\{e_j : j \in I_1\}.\]

19. \[e' = (e_1, e_2, \ldots, e_L).\]

20. \[\tilde{e}_1 = M\{\tilde{e}_j : j \in I_1\}.\]

21. \[\tilde{e}' = (\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_L).\]

22. \[Y_i = \text{median of the adjusted dependent variable in the }i\text{-th cell and }\]

\[Y_i = M\{y_j(\tilde{\beta}_1) : j \in I_i\} = X_i'\beta + \tilde{\varepsilon}_i.\]

23. \[n_i^* = \begin{cases} (n_i + 1)/2 & \text{if } n_i \text{ is odd} \\ (n_i/2) + 1 & \text{if } n_i \text{ is even.} \end{cases}\]

24. \[m_i = \begin{cases} 1 & \text{if } y_j > 0 \text{ for } j = 1, 2, \ldots, n_i^* \\ 0 & \text{otherwise,} \end{cases}\]

where \(j\) is ordered according to the adjusted dependent variable.
variable \( y_j(\beta_1) \). The median of the \( i \)-th cell is "observed", iff \( m_i = 1 \) and "unobserved", iff \( m_i = 0 \).

25. \( L_0 \): number of cells where \( m_i = 0 \).

26. \( L_1 \): number of cells where \( m_i = 1 \).

27. \( X_i^0, \beta_1 = \min_{k \in I_1} x_k^0, \beta_1 \).

28. \( X_i^1 = (1, X_i^0) \).

29. \( b_i = \lim_{N \to \infty} \inf_{j \in I_1} x_j^i \beta_i \).

30. \( \hat{\beta} \): second-stage estimator.
Appendix B

I first show \( \hat{\sigma} \) is essentially bounded. \( \hat{\sigma}^* \), defined in (3.5), is

\[
(B.1) \quad \hat{\sigma}^* = \sum_1 \frac{n_i (Y_i - X_i \hat{\beta}^*)^2}{L_1}
\]

\[
= \sum_1 n_i (\tilde{\varepsilon}_i - X_i \hat{\beta}^* \text{vec}(X_i X_i)^{-1} \text{vec} \{X_i \hat{\varepsilon}_i \})^2 / L_1.
\]

From Proposition 3.1, \( \sqrt{n_i} (\tilde{\varepsilon}_i - \varepsilon_i) \xrightarrow{p} 0 \) if \( X_i \beta > 0 \).

From the Law of Large Numbers,

\[
(B.2) \quad P[Y_i = 0 | X_i \beta > 0] \rightarrow 0,
\]

and

\[
(B.3) \quad P[Y_i > 0 | X_i \beta < 0] \rightarrow 0.
\]

Hence, \( \hat{\sigma}^* \equiv \sum_1 \frac{1}{n_i} \sum_{j \in X_i \beta > 0} \left( p_{jk} \right) (\sqrt{n_i} \varepsilon_i)^2 \).

Since \( \sqrt{n_i} \varepsilon_i \xrightarrow{N(0,1/4r(0)^2)} \) and \( \text{plim} \frac{n_i}{n_k} = \frac{p_i}{p_k} \), for any \( \varepsilon > 0 \), there exist \( M_\varepsilon > 0 \) and \( N_\varepsilon \) such that,

\[
(B.4) \quad P[|\hat{\sigma}^* - M_\varepsilon| < \varepsilon/2 \text{ for } N > N_\varepsilon],
\]

and

\[
(B.5) \quad \frac{L(\hat{\beta}^*, \hat{\sigma})}{\Pi \sqrt{n_i}} = \left[ \prod_1 \Phi(-\sqrt{n_i} X_i \hat{\beta}^* / \hat{\sigma}^*) \prod_1 \frac{1}{\sqrt{2\pi\hat{\sigma}^*}} \right] \varepsilon^{-1/2} L_1.
\]

From (3.13) \( \Pi \Phi(-\sqrt{n_i} X_i \hat{\beta}^* / \hat{\sigma}^*) \xrightarrow{p} 1 \).

Therefore, there exists \( N_1 \) such that

\[
(B.6) \quad P\left[ \frac{L(\hat{\beta}^*, \hat{\sigma}^*)}{\Pi \sqrt{n_i}} > c_0 \hat{\sigma}^* - L_1 \right] \rightarrow 1 - \varepsilon/2
\]

\[
- L_1/2
\]

where \( c_0 = (2\pi e)^{-1/2} \).
Here,

\[(B.7) \quad \sigma^{−L} \rightarrow L(\hat{\beta}, \hat{\sigma})/\Pi_{l} n_{1}\]

and

\[(B.8) \quad L(\hat{\beta}, \hat{\sigma}) > L(\beta^*, \sigma^*).\]

From (B.7) and (B.8),

\[(B.9) \quad P[\sigma^{−L} \rightarrow c_{0} \sigma^{−L}] > P[L(\hat{\beta}^*, \sigma^*)/\Pi_{l} n_{1}] > c_{0} \sigma^{−L}].\]

From (B.6) and (B.9),

\[(B.10) \quad P[\sigma > c_{0}^{−1/L} \sigma^*] < \varepsilon/2.\]

From (B.4) and (B.10) there exists \(M^*_\varepsilon\) such that,

\[(B.11) \quad P[\sigma > M^*_\varepsilon] < \varepsilon \text{ for } N > N^*_\varepsilon.\]

Hence, \(\hat{\sigma}\) is essentially bounded.

From (B.2), (B.3), and (B.11), we get,

\[(B.12) \quad \Phi(-\sqrt{n}_{1} X_{1}^{i} \hat{\beta}/\hat{\sigma}) \rightarrow 0 \quad \text{if } Y_{i} = 0,\]

and

\[(B.13) \quad \Phi(-\sqrt{n}_{1} X_{1}^{i} \hat{\beta}/\hat{\sigma}) \rightarrow 1 \quad \text{if } Y_{i} = 0.\]

From (B.11), (B.12), and (B.13), the second term of (4.12) goes to zero in probability.
Footnotes

1. Steps (i) and (ii) may be replaced by Manski's maximum score estimator (Manski [1975]) which gives a consistent estimator of $\beta/\sigma$ if $F(0)=0.5$.

2. Unless explicitly stated, equalities and inequalities involving random variables in this paper are supposed to hold every for possible realization of random variables.

3. This condition may replaced by the following.

   a) \[ \lim_{N \to \infty} \max_k \max_j \sum_{l=1}^{N} |\text{cov}(\eta_j, \eta_k)| < M_1 < \infty \], and

   b) \[ \lim_{N \to \infty} \max_{\lambda} \max_k \sum_{j=1}^{N} |\text{cov}(\phi^\lambda_j, \phi^k_k)| < M_2 < \infty \],

   for some $M_1$ and $M_2$, where $\phi^\lambda_j$ and $\eta_j$ are defined in (3.40) and (3.42). The proof under these conditions is identical to the original one. Therefore, the estimator is robust with respect to autocorrelation if the error terms are symmetrically distributed.
References


Goldberger, A.S. [1980], "Abnormal Selection Bias", SSRI Workshop Series No. 8006, University of Wisconsin


Hurd, M. [1979], "Estimation in Truncated Samples when there is Heteroscedasticity", *Journal of Econometrics*, Vol. 11, 247-258


Nawata, K. [1985], "Robust Estimation Based on Grouped-Adjusted Data in Linear Regression Models", Econometric Workshop, Technical Report No.8, Stanford University

Paarsch, H.J. [1983], "A Monte Carlo Comparison of Estimators for


Technical Reports in This Series


