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

KAZUMITSU NAWATA

TECHNICAL REPORT NO. 8
AUGUST 1985

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT NO. SES 82-08180

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

KAZUMITSU NAWATA

Technical Report No. 8
August 1985

PREPARED UNDER
NATIONAL SCIENCE FOUNDATION GRANT NO. SES 82-08180

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

By
Kazumitsu Nawata*

1. Introduction

To estimate linear regression models, the ordinary least squares (OLS) method is usually used. Although the OLS estimator is computationally attractive and the best unbiased estimator when the distribution of error terms is normal, it does not satisfy a robustness requirement; that is, it is sensitive to the distribution of the error terms. Especially when the error terms have a fat-tailed distribution, poor performance of the OLS estimator is well known. Since the OLS minimizes squared deviations, it gives relatively heavy weight to extreme values. Hence, it is sensitive to extreme values. As many researchers (e.g. Mandebrot [1969], and Fama [1970]) have pointed out, many economic data series are not normal and follow fat-tailed distributions such as the Cauchy distribution. In these cases robust estimators, such as the least absolute deviation (LAD) estimator, are superior estimators.

The LAD estimator minimizes \( \sum |y_j - x_j'\beta| \). It is well known that the

*This work was supported by National Science Foundation Grant SES 82-08180 at the Department of Economics, Stanford University. I would like to thank my principal adviser, Takeshi Amemiya, for numerous helpful suggestions through several previous drafts. I would like to also thank Theodore W. Anderson, Timothy F. Bresnahan and Lawrence J. Lau for their helpful comments.
LAD estimator is a superior robust estimator when the error terms have a fat-tailed distribution. However, since calculation of the LAD estimator is computationally burdensome, it has not been widely used in the estimation of economic models.

In this paper a robust estimator based on grouped, adjusted data is proposed. The estimator is

i) computationally feasible,
ii) consistent under general assumptions, and
iii) asymptotically as efficient as the LAD estimator.

The estimator can also be used to estimate censored regression models (Nawata [1985]).

The estimator is based on grouping and adjustment. First, the proper domain of the independent variables is divided into a finite number of cells. Then the median of the dependent variable in each cell is estimated. The next-round estimator is calculated by OLS using the parameters estimated in the previous round. These operations are continued until the procedure converges. The final estimator is calculated by WLS based on the data adjusted by the previous-stage estimator.

In Chapter 2, the estimator is defined. In Chapter 3, the estimator is shown to be consistent. The asymptotic properties of the estimator are studied in Chapter 4. Results of a Monte Carlo study are in given Chapter 5. The symbols used in the paper are summarized in Appendix A.
2. Model and Definition of the Estimator

2.1 Model

The model studied in this paper is

\begin{align}
  y_j &= \beta_0 + x_j^0 \beta_1 + u_j \\
  &= x_j^\beta + u_j, \quad j=1,2,3,\ldots,N,
\end{align}

where \( u_j \) are random variables with median 0 and \( x_j^0 \) are random variables

which satisfy the following assumptions. Additional assumptions on

\{u_j\} and \{x_j^0\} will be stated whenever they are needed to prove

subsequent theorems.

[Assumption 2.1]

i) The quasi-support \( \Omega \) of \( x_j^0 \) is fixed and bounded.

ii) Every value of \( x_j^0 \) belongs to \( \Omega \).

iii) There is no perfect multicollinearity among elements of

the vector \( x_j^0 \).

The quasi-support is defined in Definition 2.1.

[Definition 2.1]

A point of the quasi-support is a point such that the number of

observations in any open neighborhood is \( O_p(N) \) where \( N \) is the number of

total observations and a random variable \( \xi \) is \( O_p(N) \) if and only if

\( \xi/N = O_p(1) \) and \( N/\xi = O_p(1) \). The quasi-support is the set of all such

points. If \( x_j^0 \) are i.i.d., the quasi-support is the same as the support

of their common distribution function.
Other types of models can be also estimated by the proposed method. These models are described in Appendix B.

2.2 Definition of the Estimator

The estimation procedure consists of two different stages.

The first stage consists of the following steps.

i) Divide the sample space of independent variables into a finite number of cells.

ii) Estimate the median of the dependent variable in each cell.

iii) Using (ii), calculate the i-th round estimator, \( \beta_1^i \), by ordinary least squares.

iv) Using \( \beta_1^i \), adjust the value of the observations.

v) Repeat steps (ii) through (iv).

vi) Define the first-stage estimator \( \bar{\beta}_1 \) as

\[
\bar{\beta}_1 = \begin{cases} 
\lim_{i \to \infty} \beta_1^i & \text{if all the cells are nonempty} \\
0 & \text{otherwise.}
\end{cases}
\]

The purpose of this stage is to obtain a consistent estimator of order \( N^{-1/2} \), which is then used to calculate the second-stage estimator.

As shown in Chapter 4, the asymptotic properties of the second-stage estimator do not depend on the first-stage estimator as long as it is a consistent estimator of order \( N^{-1/2} \).

The second-stage estimator is then calculated by weighted least
squares using the data which have been adjusted by the first-stage estimator. The asymptotic properties of the second-stage estimator are studied in Chapter 4.

In this paper the median of the sequence of real numbers \( \{a_j\} \) is defined as

\[
M[\{a_j\}] = \begin{cases} 
  a_n & \text{if } n \text{ is odd} \\
  \frac{(a_{n_2} + a_{n_3})}{2} & \text{if } n \text{ is even},
\end{cases}
\]

where \( n \) is a number of elements in \( \{a_j\} \),
\[
a_1 \succ a_2 \succ \ldots \succ a_n,
\]
\[
n_1 = (n + 1)/2, \quad n_2 = n/2, \quad \text{and } n_3 = n_2 + 1.
\]

2.2.1 Definition of the First-Stage Estimator

To estimate the first-stage, the sample space of the independent variables is divided into \( L+1 \) nonoverlapping cells \( S_1, S_2, \ldots, S_L, \emptyset \).

\( S_1, S_2, \ldots, S_L, \emptyset \) and fixed \( K \)-dimensional vectors, \( X_1^o, X_2^o, \ldots, X_L^o \), are taken so that the following conditions are satisfied.

i) \( S_i \) is fixed,

ii) \( S_i \) is bounded,

iii) \( S_i \) is convex,

iv) \( S_i \) is open, and \( S_i \cap \emptyset \neq \emptyset \), \( i = 1, 2, \ldots, L \).

v) \( S = (\bigcup_{i=1}^L S_i)^c \)

vi) \( X_i^o \in S_i, i = 1, 2, \ldots, L \).

vii) \( X^o\pi^o \) is a nonsingular matrix

where \( X_i^o = x_i^o / x_i^o \) is defined as \( X_i^o = (x_i^o, x_i^o, \ldots, x_i^o) \).
\begin{align*}
\text{viii). Max } & \sum_{k=1}^{K} \sum_{j=1}^{L} |a_{ji}|d_{ik} < 1 \\
\text{where } a_{ji} \text{ is the } (j,i)\text{th element of } (X^O',X^O)^{-1}X^O, \\
d_{ik} &= \sup_{v \in I_i} |\text{the } k\text{-th element of } (x^O_v-x^O_i)|, \text{ and} \\
& v \in I_i \text{ if and only if } x^O_v \in S_i.
\end{align*}

As shown in Theorem 2.1, \(S_1, S_2, \ldots, S_L\) and \(X^O_1, X^O_2, \ldots, X^O_L\) which satisfy these conditions always exist. The first-stage estimator is estimated by the observations which belong to \(U \cup S_i\).

If the \(i\)-th cell is nonempty, \(M[A_i(\alpha)] = M\{y_j-(x^O_j-x^O_i)'\alpha|j \in I_i}\) is well defined. Let

\begin{equation}
\mu_i(\alpha) = \begin{cases} 
M[A_i(\alpha)] & \text{if the } i\text{-th cell is nonempty} \\
0 & \text{otherwise},
\end{cases}
\end{equation}

and

\begin{equation}
\mu'(\alpha) = (\mu_1(\alpha), \mu_2(\alpha), \ldots, \mu_L(\alpha)).
\end{equation}

Define the mapping \(T(\alpha)\) as \(2/\)

\begin{equation}
T(\alpha) = (X^O',X^O)^{-1}X^O'\mu(\alpha).
\end{equation}

As is shown in Theorem 3.3, \(T(\alpha)\) defined here is a contraction mapping and therefore has a unique fixed point.

The first-stage estimator \(\bar{\beta}_1\) is defined as

\begin{equation}
\bar{\beta}_1 = \begin{cases} 
a \text{ fixed point of } T(\alpha) & \text{if all the cells are nonempty} \\
0 & \text{otherwise.}
\end{cases}
\end{equation}
The following theorem proves the existence of \( S_1, S_2, \ldots, S_L \) and \( X_1^0, X_2^0, \ldots, X_L^0 \) which satisfy (i) through (viii).

|Theorem 2.1|

Under |Assumption 2.1|, \( S_1, S_2, \ldots, S_L \) and \( X_1^0, X_2^0, \ldots, X_L^0 \) always exist which satisfy (i) through (viii).

|Proof|

If there is no perfect multicollinearity among \( x_j^0 \), it is always possible to take \( X_1^0, X_2^0, \ldots, X_L^0 \) so that \( X^0 \) is nonsingular.

Let \( S_i = \{z | \rho(z, x_i^0) < \delta_i \} \) where \( \rho \) is a metric and \( \delta_i \) is a constant.

It is clear that \( S_i \cap \Omega \neq \emptyset \) and \( d_{ik} \to 0 \) \( k = 1, 2, \ldots, K \), as \( \delta_i \to 0 \). Here,

\[
(2.7) \quad \sum_{j=1}^{K} \sum_{i=1}^{L} |a_{ji}| d_{ik} \to 0 \text{ as } d_{ik} \to 0 \text{ for all } i \text{ and } k.
\]

Hence, if we make \( \delta_i \) small enough for all \( i \), we get

\[
(2.8) \quad \max_{k} \sum_{j=1}^{K} \sum_{i=1}^{L} |a_{ji}| d_{ik} < 1.
\]

Hence, \( S_1, S_2, \ldots, S_L \) and \( X_1^0, X_2^0, \ldots, X_L^0 \) always exist which satisfy the required conditions.

QED.

The next corollary gives an example of \( S_1, S_2, \ldots, S_L \) and \( X_1^0, X_2^0, \ldots, X_L^0 \).

|Corollary 2.1|
Let \( a_1, a_2, \ldots, a_L \) and \( b_1, b_2, \ldots, b_L \) be sequences of constant numbers which satisfy \( a_k < b_k \) for all \( k \). Let \( v \) be an integer. Define

\[
\Psi(i_1, i_2, \ldots, i_K) = \{ z | a_k + \frac{i_k - 1}{v} c_k < z < a_k + \frac{i_k}{v} c_k, k = 1, 2, \ldots, K \},
\]

where \( z \) is a \( K \)-dimensional vector,
\( z_k \) is the \( k \)-th element of \( z \), and
\[
c_k = b_k - a_k.
\]

Let \( x^0(i_1, i_2, \ldots, i_K) \) be the \( K \)-th dimensional vector whose \( k \)-th element is \( a_k + c_k(i_k - 0.5)/v \). Then there exists \( v*(K) \) such that \( \{ \Psi(i_1, i_2, \ldots, i_K) \} \) and \( \{ x^0(i_1, i_2, \ldots, i_K) \} \) satisfy all the required conditions except (iii) for \( v \approx v*(K) \). Therefore, if each cell contains at least one point of quasi-support in its interior, all the conditions are satisfied.

[Proof]

We can normalize \( x^0_j \) by dividing the \( k \)-th element by \( w_k = c_k/2v \). Therefore, without loss of generality, we can assume that the length of each interval is 2. In this case, \( L = v^K \). Let \( X^0_1 \) be the center of the \( i \)-th cell. Then under these conditions,

\[
x^0, \hat{x}^0 = v^{K-1} \delta I
\]

where

\[
\delta = v^{k-1}(v^3-v)/3.
\]

Therefore,

\[
a_{ji} = x_{ji}/\delta \text{ where } x_{ji} \text{ is the } (j,i)\text{-th element of } \hat{x}^0.
\]

Since the length of each interval is 2, \( d_{ik} = 1 \) for all \( i \) and \( k \). Therefore,
\[(2.13) \quad \sum_{j=1}^{K} \sum_{i=1}^{L} a_{ji} d_{ik} = \sum_{j=1}^{K} \sum_{i=1}^{L} |x_{ji}|/\delta.\]

Here \(\sum_{j=1}^{L} |x_{ji}|\) is

i) \(\nu^{K+1}/2\) if \(K\) is even, and

ii) \(\nu^{K-1}(\nu-1)(\nu+1)/2\) if \(K\) is odd.

Therefore,

\[(2.14) \quad \sum_{j=1}^{K} \sum_{i=1}^{L} a_{ji} |d_{ik}| < \eta(v, K)\]

where \(\eta(v, K)\) equals

i) \(3K\nu/2(\nu^2-1)\) if \(K\) is even, and

ii) \(3K/2\nu\) if \(K\) is odd.

Since \(\eta(v, K) \to 0\) as \(v \to 0\), there exists \(v^*(K)\) such that,

\[(2.15) \quad \max_{k} \sum_{j=1}^{K} \sum_{i=1}^{L} a_{ji} |d_{ik}| < 1 \quad \text{for} \quad v > v^*(K).\]

QED.

Note that \(v^*(K)\) is smaller than \(\min\{v | \eta(v, K) < 1\}\). \(v^*(K)\) can be calculated by the linear programming technique. Some values of \(v^*(K)\) are given as follows:

\[
\begin{array}{cccccc}
K & 1 & 2 & 3 & 4 & 5 \\
v^*(K) & 2 & 2 & 3 & 3 & 4.
\end{array}
\]
2.2.2 Definition of the Second-Stage Estimator \( \hat{\beta} \)

The second-stage estimator is calculated based on the data adjusted by the first-stage estimator \( \hat{\beta}_1 \). Let \( S_1, S_2, \ldots, S_{L^*} \) be nonempty, convex, bounded cells which satisfy

\[ P[\text{the smallest characteristic root of } X'X < \delta^*] \rightarrow 0 \quad \text{for some } \delta^* > 0. \]

Under Assumption 2.1, it is always possible to take cells so that this condition is satisfied. \( S_1, S_2, \ldots, S_{L^*} \) do not have to satisfy the conditions of the first-stage estimator and \( L^* \) does not have to equal \( L \).

In this stage, the value of the \( x \)-variable which represents the \( i \)-th cell must be the sample mean of \( x_j^O \) in the cell. Namely,

\[ (2.16) \quad x_i^O = \sum_{j \in I_i} x_j^O / n_i. \]

Define

\[ (2.17) \quad X' = (X_1', X_2', \ldots, X_{L^*}') \quad \text{where } X_1' = \begin{pmatrix} 1, x_1^O' \end{pmatrix}. \]

and

\[ (2.18) \quad \Sigma = \begin{bmatrix} n_1/N & 0 & \cdots & 0 \\ 0 & n_2/N & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_{L^*}/N \end{bmatrix}. \]

Then the second-stage estimator is

\[ (2.19) \quad \hat{\beta} = (X' \Sigma X)^{-1} X' \Sigma \mu'(\hat{\beta}_1) \]

where \( \mu'(\hat{\beta}_1) = (M[A_1(\hat{\beta}_1)], M[A_2(\hat{\beta}_1)], \ldots, M[A_{L^*}(\hat{\beta}_1)]). \)
3. Consistency of the Estimator

In this section, the first-stage estimator $\bar{\beta}_1$ and the second-stage estimator $\hat{\beta}$ are shown to be consistent.

3.1 Consistency of the First-Stage Estimator $\bar{\beta}_1$

In this section the first-stage estimator $\bar{\beta}_1$ is shown to be consistent. Since the properties of the contraction mapping and fixed-point theorems are used in the proof, these theorems are presented before the proof of consistency.

3.1.1 Contraction Mapping and Fixed Point Theorems

[Definition 3.1]

Let $\Psi$ be a mapping of a metric space $A$ into itself. $\Psi$ is a contraction mapping if there exists $\gamma < 1$ such that

$$\rho(\Psi(\alpha_1), \Psi(\alpha_2)) \leq \gamma \rho(\alpha_1, \alpha_2)$$

for any $\alpha_1, \alpha_2 \in A$ where $\rho$ is a metric.

[Theorem 3.1]

Every contraction mapping defined on a complete metric space has a unique fixed point.

[Proof] See Kolmogorov and Fomin [1975]

[Theorem 3.2]

Let $\bar{\alpha}$ be the unique fixed point of a contraction mapping $\Psi$.

Then $\lim \alpha_n = \bar{\alpha}$ where $\alpha_{n+1} = \Psi(\alpha_n)$.

[Proof]
Since \( \Psi \) is a contraction mapping,

\[
(3.1) \quad \rho(\alpha_{n+1}, \bar{\alpha}) = \rho(\Psi(\alpha_n), \Psi(\bar{\alpha})) < \gamma \rho(\alpha_n, \bar{\alpha}).
\]

Therefore,

\[
(3.2) \quad \rho(\alpha_n, \bar{\alpha}) < \gamma^n \rho(\alpha_0, \bar{\alpha}).
\]

Since \( |\gamma| < 1 \),

\[
(3.3) \quad \lim_{n \to \infty} \rho(\alpha_n, \bar{\alpha}) = 0.
\]

Hence,

\[
(3.4) \quad \lim_{n \to \infty} \alpha_n = \bar{\alpha}.
\]

QED.

3.1.2 Consistency of \( \bar{\beta}_1 \)

First it is shown in Theorem 3.3 that \( T(\alpha) \) defined in (2.5) is a contraction mapping. Consistency of the first stage estimator is proved in Theorem 3.4.

[Theorem 3.3]

\( T(\alpha) \) is a contraction mapping.

To prove Theorem 3.3, the following proposition is used.

[Proposition 3.1]

\[
(3.5) \quad |\mu_i(\alpha) - \mu_i(\alpha^*)| < d_i' |\alpha - \alpha^*|
\]

where \( d_i \) is a \( K \) dimensional vector whose \( k \)-th element...
is sup \(|\text{the } k\text{-th element of } (x_j^0 - x_i^0)|,\)
\(j \in \mathcal{I}_1^0,\)
\(|\alpha - \alpha^*| = (|\alpha_1 - \alpha_1^*|, |\alpha_2 - \alpha_2^*|, \ldots, |\alpha_K - \alpha_K^*|),\) and
\(\mu_1(\alpha) = \begin{cases} M|A_i(\alpha)| & \text{if the } i\text{-th cell is nonempty} \\ 0 & \text{otherwise.} \end{cases}\)

[Proof]

If the \(i\)-th cell is empty, the proof is trivial. Suppose the \(i\)-th cell is nonempty. Then for \(j \in \mathcal{I}_1^0,\)
\begin{equation}
(3.6) \quad y_j(\alpha) = y_j - (x_j^0 - x_i^0)'\alpha.
\end{equation}
Hence,
\begin{equation}
(3.7) \quad y_j(\alpha) = y_j(\alpha^*) + (x_j^0 - x_i^0)'(\alpha - \alpha^*).
\end{equation}
Therefore,
\begin{equation}
(3.8) \quad y_j(\alpha^*) - |(x_j^0 - x_i^0)'(\alpha - \alpha^*)| < y_j(\alpha) < y_j(\alpha^*) + |(x_j^0 - x_i^0)'(\alpha - \alpha^*)|.
\end{equation}
Let \(\delta_i = |\alpha - \alpha^*|\). Since \(\delta_i > |(x_j^0 - x_i^0)'(\alpha - \alpha^*)|,\)
\begin{equation}
(3.9) \quad y_j(\alpha^*) - \delta_i < y_j(\alpha) < y_j(\alpha^*) + \delta_i.
\end{equation}
From (2.2) and (3.9),
\begin{equation}
(3.10) \quad M\{y_j(\alpha^*) - \delta_i\} < M\{y_j(\alpha)\} < M\{y_j(\alpha^*) + \delta_i\}.
\end{equation}
From (2.2) and (3.10),
\begin{equation}
(3.11) \quad M|A_i(\alpha^*)| - \delta_i < M|A_i(\alpha)| < M|A_i(\alpha^*)| + \delta_i.
\end{equation}
Therefore,
\begin{equation}
(3.12) \quad |\mu_1(\alpha) - \mu_1(\alpha^*)| = |M|A_i(\alpha)| - M|A_i(\alpha^*)|| < \delta_i.
\end{equation}
QED.

Next, Theorem 3.3 is proved.

[Proof of Theorem 3.3]
Define the metric

\begin{equation}
\rho(\alpha, \alpha^*) = \sum_{i=1}^{K} |\alpha_k - \alpha_k^*|
\end{equation}

where \( \alpha' = (\alpha_1, \alpha_2, \ldots, \alpha_K) \) and \( \alpha^*' = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_K^*) \).

Then

\begin{equation}
\rho(T(\alpha), T(\alpha^*)) = \sum_{j=1}^{K} \sum_{i=1}^{L} a_{ij} |\mu_i(\alpha) - \mu_i(\alpha^*)|
\end{equation}

where \( a_{ij} \) is the \((i,j)\)-th element of \((X^0, X^0)'\).

Here,

\begin{equation}
\rho(T(\alpha), T(\alpha^*)) < \sum_{j=1}^{K} \sum_{i=1}^{L} |a_{ij}| |\mu_i(\alpha) - \mu_i(\alpha^*)|
\end{equation}

\begin{align*}
&= \sum_{j=1}^{K} \sum_{i=1}^{L} |a_{ij}| \left( \sum_{k=1}^{K} d_{ik} |\alpha_k - \alpha_k^*| \right) \\
&= \sum_{j=1}^{K} \sum_{i=1}^{L} |a_{ij}| \left( \sup_{j \in I_1} \left( \sum_{k=1}^{K} d_{ik} |\alpha_k - \alpha_k^*| \right) \right)
\end{align*}

\begin{align*}
&= \gamma_o \rho(\alpha, \alpha^*),
\end{align*}

where \( d_{ik} = \sup_{j \in I_1} \left| (\text{the } k\text{-th element of } (x_j^0 - x_i^0)) \right| \),

and \( \gamma_o = \max \left\{ \sum_{k=1}^{K} \sum_{i=1}^{L} |a_{ij}| d_{jk} \right\} < 1 \).

The second inequality is given by Proposition 3.1. The last inequality is given by the definition of the grouping rule.

Hence \( T(\alpha) \) is a contraction mapping.

QED.
The following theorem shows that \( \bar{\beta}_1 \), defined in (2.20) is a consistent estimator of \( \beta_1 \).

[Theorem 3.4]

Let \( \lim_{N \to \infty} M[\{u_j | j \in I_1\}] = 0 \) for all \( i \). Then \( \bar{\beta}_1 \) is a consistent estimator of \( \beta_1 \).

Note that the above assumption is implied by separate assumptions on \( u_j \) and \( x_j^0 \). For example, the assumptions that \( \{u_j\} \) and \( \{x_j^0\} \) are independent and \( \lim_{N \to \infty} M[\{u_j\}] = 0 \) together with Assumption 2.1 imply this condition. Further, \( \lim_{N \to \infty} M[\{u_j\}] \) is implied by the assumption that \( u_j \) are independent and have positive density in the neighborhood of zero.

[Proof]

Since \( P[\text{there exists a nonempty cell}] \) goes to zero as \( N \) goes to infinity, it is enough to show the case where all the cells are nonempty. Suppose all the cells are nonempty. Then for \( j \in I_1 \),

\[
y_j(\beta_1) = y_j - (x_j^0 - x_{j1}^0)' \beta_1
= \beta_0 + x_j^0 \beta_1 + u_j.
\]

Let \( \theta_1 = M[\{u_j | j \in I_1\}] \). Then

\[
M[A_1(\beta_1)] = M[\{y_j(\beta_1) | j \in I_1\}]
= M[\{\beta_0 + x_{j1}^0 \beta_1 + u_j | j \in I_1\}]
= \beta_0 + x_{j1}^0 \beta_1 + \theta_1.
\]

Hence,

\[
T(\beta_1) = (x^0, x^0)^{-1} x^0 \mu(\beta_1)
\]
\[ \hat{\beta} = \beta_1 + (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top \theta \]

where \( \theta' = (\theta_1, \theta_2, \ldots, \theta_L) \).

Since \( T \) is a contraction mapping,

\[
(3.19) \quad \rho(\beta_1, \bar{\beta}_1) < \rho(\beta_1, T(\beta_1)) + \rho(T(\beta_1), \bar{\beta}_1) \\
= \rho(\beta_1, T(\beta_1)) + \rho(T(\beta_1), T(\bar{\beta}_1)) \\
< \rho(\beta_1, T(\beta_1)) + \gamma \rho(\beta_1, \bar{\beta}_1)
\]

where \( 0 < \gamma < 1 \).

Therefore,

\[
(3.20) \quad \rho(\beta_1, \bar{\beta}_1) < \rho(\beta_1, T(\beta_1))/(1-\gamma).
\]

From (3.18)

\[
(3.21) \quad \rho(T(\beta_1), \beta_1) = \rho[(\hat{X}^\top \hat{X})^{-1} \hat{X}^\top \theta, 0]
\]

Since \( \operatorname{plim} \{u_j | j \in I_i\} = 0 \) for all \( i \), \( \operatorname{plim} (\hat{X}^\top \hat{X})^{-1} \hat{X}^\top \theta = 0 \).

Therefore,

\[
(3.22) \quad \rho(T(\beta_1), \beta_1) \to 0 \quad \text{in probability.}
\]

From (3.20) and (3.22),

\[
(3.23) \quad \rho(\beta_1, \bar{\beta}_1) \to 0 \quad \text{in probability.}
\]

Hence, \( \bar{\beta}_1 \) is a consistent estimator of \( \beta_1 \).

QED.

3.2 Consistency of the Second-Stage Estimator \( \hat{\beta} \)

Next it is shown that the second-stage estimator \( \hat{\beta} \) is a consistent estimator. \( \hat{\beta} \) is defined as

\[
(3.24) \quad \hat{\beta} = (X'X)^{-1}X'\Sigma u(\bar{\beta}_1).
\]

As in the definition in Section 2.3.4, the x-variables which represent each cell are the mean of the variables in each cell and the X matrix
contains the vector of 1's. The proof of the consistency is in the next theorem.

(Theorem 3.5)

The second stage estimator \( \hat{\beta} \) is a consistent estimator if

\[
\lim_{N \to \infty} M\{u_j | j \epsilon I_1\} = 0 \quad \text{for all } i.
\]

[Proof]

For \( j \epsilon I_1 \),

\[
y_j \left( \bar{\beta}_1 \right) = y_j - (x_j^O - x_1^O)' \bar{\beta}_1 = \beta_0 + x_1^O \beta_1 - (x_j^O - x_1^O)'(\bar{\beta}_1 - \beta_1) + u_j = x_1^O \beta - (x_j^O - x_1^O)'(\bar{\beta}_1 - \beta_1) + u_j.
\]

Since the size of the i-th cell is finite, there exists \( H > 0 \) such that

\[
x_1^O \beta - H \rho(\bar{\beta}_1, \beta_1) + u_j < y(\bar{\beta}_1) < x_1^O \beta + H \rho(\bar{\beta}_1, \beta_1) + u_j.
\]

Hence,

\[
|M |A_i(\bar{\beta}_1)| - X_1^O \beta| < H \rho(\bar{\beta}_1, \beta_1) + |M |\{u_j | j \epsilon I_1\}||.
\]

Since \( \rho(\bar{\beta}_1, \beta_1) \to 0 \) and \( M |\{u_j | j \epsilon I_1\}| \to 0 \) for all \( i \),

\[
P[M |A_i(\bar{\beta}_1)| - X_1^O \beta] \to P
\]

From the assumption, the smallest characteristic root of \( X' \Sigma X \) is bounded away from zero with probability approaching one. Therefore,

\[
\hat{\beta} \to P \beta.
\]

QED.
4. Asymptotic Distribution of The Estimator

This section discusses the asymptotic distribution of the estimator. It is shown that

\[(4.1) \quad \sqrt{N} (\hat{\beta} - \beta) = N(0, \frac{1}{4f(0)^2} \text{plim}\ (X' \Sigma X)^{-1}).\]

The following assumptions are made in this section.

\[(4.2) \quad u_j \text{ are i.i.d. with median 0.}\]

\[(4.3) \quad u_j \text{ have a density function } f(u). \text{ In the neighborhood of } u=0, f(u) \text{ satisfies:}\]

a) continuous and differentiable,

b) \(f(u)>0,\)

c) \(f'(u)\) is bounded.

\[(4.4) \quad x_j^0 \text{ are i.i.d. and bounded.}\]

\[(4.5) \quad \{u_j\} \text{ and } \{x_j^0\} \text{ are independent.}\]

\[(4.6) \quad s_i \text{ is fixed for all } i.\]

To prove the asymptotic normality of the estimator the following three theorems are used.

[Theorem 4.1]

Suppose \(u_j\) satisfies \(4.2\) and \(4.3\).

Then

\[(4.7) \quad \sqrt{\tau} \text{M}\{u_j\} + N(0, \frac{1}{4f(0)^2})\]
where \( \tau \) is the number of elements in \( \{u_j\} \).

[Proof] See, for example, Amemiya [1985, p.150].

[Theorem 4.2]

Let

\[
\hat{\theta}_\tau = M[\{u_j\}],
\]

and

\[
\bar{\theta}_\tau = M[\{u_j + v_j w_\tau\}]
\]

where \( \tau \) is the number of elements in \( \{u_j\} \), and \( v_j \) and \( w_\tau \) are \( K \)-dimensional vectors.

Suppose \( u_j, v_j, \) and \( w_\tau \) satisfy the following assumptions.

i) \( u_j \) satisfy (4.2) and (4.3).

ii) \( v_j \) are i.i.d. and bounded.

iii) \( u_j \) and \( v_j \) are independent for all \( j \).

iv) \( (u_j, v_j) \) and \( (u_k, v_k) \) are independent for all \( j \neq k \).

v) For any \( \varepsilon > 0 \), there exist \( M_\varepsilon \) and \( \tau_0 \) such that

\[
P[\rho(\sqrt{\tau} w_\tau, 0) > M_\varepsilon ] < \varepsilon \text{ for } \tau > \tau_0 .
\]

Then

\[
\sqrt{\tau}(\hat{\theta}_\tau - \bar{\theta}_\tau) \to 0 \text{ in probability.}
\]

The proof is in Appendix C. The outline of the proof is as follows:

i) Count the number of \( u_j \) which satisfy

\[
u_j < \hat{\theta}_\tau \text{ and } u_j + v_j w_\tau > \hat{\theta}_\tau .
\]

ii) Count the number of \( u_j \) which satisfy
\[ u_j > \hat{\theta}^\tau \text{ and } u_j + v_j' w < \hat{\theta}^\tau. \]

iii) Let D be the difference between (i) and (ii).

Then D is $o_P(\sqrt{\tau})$.

iv) If $\sqrt{\tau}(\hat{\theta}^\tau - \theta^\tau) > \varepsilon$, plim $D/\sqrt{\tau} > \delta$ for some $\delta > 0$.

v) Therefore $\sqrt{\tau}(\hat{\theta}^\tau - \theta^\tau) \to 0$.

[Theorem 4.3]

Under the assumptions of (4.3)-(4.6), for any $\varepsilon > 0$ there exist $M_\varepsilon$ and $N_0$ such that

\[ P[\rho(\sqrt{N}(\hat{\theta}^\tau - \beta^\tau)_t, 0) > M_\varepsilon] < \varepsilon \quad \text{for } N > N_0 \]

where $\hat{\beta}^\tau$ is the first-stage estimator.

[Proof]

Since $P[\text{there exists a nonempty cell}]$ goes to zero as \( N \) goes to infinity, it is enough to show the case where all the cells are nonempty. If all the cells are nonempty:

\[ T(\beta^\tau) = \beta^\tau + (\chi^\tau \chi^\tau)^{-1} \chi^\tau \theta \]

where $\theta = (\theta_1, \theta_2, \ldots, \theta_L)$ and $\theta^\tau = M[(u_j | j \in I_\tau)]$.

Since the $\chi^\tau_j$ are i.i.d., plim $n_1/N = P_1$ = constant. Therefore, from Theorem 4.1, $\sqrt{n_1} \theta_\tau$ converges to a normal random variable with mean zero and finite variance and

\[ \sqrt{N} \theta_\tau \to \text{normal distribution with mean zero and finite variance.} \]

Since $\theta_\tau$ are independent,

\[ \sqrt{N} [T(\beta^\tau) - \beta^\tau] \to \text{normal distribution with mean zero and finite variance.} \]
Therefore, for any \( \varepsilon, \delta > 0 \), there exists \( M_{\varepsilon} \) and \( N_{1} \) such that
\[
(4.15) \quad P\{ \rho (\sqrt{N}(T(\beta_1) - \beta_1), 0) > \delta M_{\varepsilon} \} < \varepsilon / 2 \quad \text{for } N > N_{1}.
\]
Here \( T \) is a contraction mapping and the first-stage estimator \( \tilde{\beta}_1 \) is its fixed point,
\[
(4.16) \quad \rho (\tilde{\beta}_1, \beta_1) < \rho (T(\beta_1), \beta_1)/(1 - \gamma) \quad \text{where } 0 < \gamma < 1,
\]
and
\[
(4.17) \quad \rho (\sqrt{N} (\tilde{\beta}_1 - \beta_1), 0) < \rho (\sqrt{N} (T(\beta_1) - \beta_1)/(1 - \gamma), 0).
\]
Hence,
\[
(4.18) \quad P\{ \rho (\sqrt{N} (\tilde{\beta}_1 - \beta_1), 0) > M_{\varepsilon} \} < P\{ \rho (\sqrt{N} (T(\beta_1) - \beta_1)/(1 - \gamma), 0) > (1 - \gamma) M_{\varepsilon} \}.
\]
From (4.15) and (4.18),
\[
(4.19) \quad P\{ \rho (\sqrt{N} (\tilde{\beta}_1 - \beta_1), 0) > M_{\varepsilon} \} < \varepsilon / 2 \quad \text{for } N > N_{1}.
\]
Therefore,
\[
(4.20) \quad P\{ \rho (\sqrt{N} (\tilde{\beta}_1 - \beta_1), 0) > M_{\varepsilon} \} < \varepsilon \quad \text{for } N > N_{0}.
\]
QED.

The following theorem shows the asymptotic normality of the second-stage estimator \( \hat{\beta} \).

[Theorem 4.4]

Under the assumptions of (4.2)-(4.6)
\[
(4.21) \quad \sqrt{N}(\hat{\beta} - \beta) \rightarrow N(0, \frac{1}{4f(0)^2} A^{-1})
\]

where \( A = \text{plim} X'X \).

[Proof]

Let \( x_{i}^* = E[x_{j}^0|j \neq i] \) and \( X_{i}^* = (1, x_{i}^*) \) and define
\[(4.22) \quad \mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_L^*)\]
\[
(4.23) \quad m_i = M[A_i(\bar{\beta}_1)] - x_i^* \bar{\beta} = M[(u_j + (x_j^0 - x_i^0)'(\bar{\beta}_1 - \bar{\beta}_1))]
\]

and
\[
(4.24) \quad \theta_i = M\{u_j | j \in I_i\}.
\]

Next, the samples in the \(i\)-th cell are considered. Let \(v_j^i = x_j^0 - x_i^0\), where \(j \in I_i\). Then \(v_j^i\) are i.i.d. with mean zero and bounded. Here, since \(x_j^0\) are i.i.d., \(\text{plim} \frac{n_i}{N} = p_i = \text{constant}\). Hence from Theorems 4.2 and 4.3,

\[
(4.25) \quad \sqrt{N}(m_i - \theta_i) \xrightarrow{p} 0 \quad \text{for all } i.
\]

Therefore,
\[
(4.26) \quad \sqrt{N}(m - \theta) \xrightarrow{p} 0
\]

where \(m = (m_1, m_2, \ldots, m_L^*)\) and \(\theta = (\theta_1, \theta_2, \ldots, \theta_L^*)\).

From Theorem 4.1,
\[
(4.27) \quad \sqrt{N} \theta_i \xrightarrow{d} N(0, p_i/(4f(0))^2).
\]

Since the \(\theta_i\)'s are independent,
\[
(4.28) \quad \sqrt{N} \theta \xrightarrow{d} N(0, (1/4f(0)^2) \text{ plim } \Sigma^{-1}).
\]

For \(j \in I_i\),
\[
(4.29) \quad y_j(\bar{\beta}_1) = y_j^0 - (x_j^0 - x_i^0)'\bar{\beta}_1 = y_j^0 + x_j^0'\bar{\beta}_1 + (x_j^0 - x_i^0)'(\bar{\beta}_1 - \bar{\beta}_1) + u_j = x_i^0'\beta + (x_i^0 - x_i^0)'(\bar{\beta}_1 - \bar{\beta}_1) + (x_j^0 - x_i^0)'(\bar{\beta}_1 - \bar{\beta}_1) + u_j.
\]

Let
\[
(4.30) \quad m_i^* = M\{y_j^0 - (x_j^0 - x_i^0) | j \in I_i\} - x_i^0'\beta.
\]

From (4.29) and (4.30),
\[
(4.31) \quad m_i^* = m_i + (x_i^0 - x_i^0)'(\bar{\beta}_1 - \bar{\beta}_1).
\]
Consequently,

\[(4.32) \quad \sqrt{N} (\hat{\beta} - \beta) = (X'\Sigma X)^{-1}X'\Sigma \sqrt{N} m + (X'\Sigma X)^{-1}X'\Sigma \sqrt{N} \phi\]

where \(\phi = (\phi_1, \phi_2, \ldots, \phi_{\ell*})\) and \(\phi_1 = (x_i^* - x_i^0)'(\beta_1 - \hat{\beta}_1)\).

From Theorem 4.3, for any \(\varepsilon > 0\) there exists \(M_\varepsilon\) such that

\[(4.33) \quad P[\rho(\sqrt{N} (\hat{\beta}_1 - \beta_1), 0) > M_\varepsilon] < \varepsilon.\]

Since \(X_i^0 = \sum_{j \in I_i} x_j^0/n_i\),

\[(4.34) \quad \text{plim} (X_i^0 - x_i^*) = 0 \quad \text{for all } i.\]

Therefore, the second term of (4.32) goes to zero in probability as \(N \to \infty\).

Hence,

\[(4.35) \quad \sqrt{N} (\hat{\beta} - \beta) = (X'\Sigma X)^{-1}X'\Sigma \sqrt{N} m\]

LD

\[= (X^*\Sigma X^*)^{-1}X^*\Sigma \sqrt{N} \Theta\]

\[\to N(0, \frac{1}{4f(0)^2} A^{-1})\]

where \(A = \text{plim} X^*\Sigma X^* = \text{plim} X^*\Sigma X\).

\[\text{QED.}\]
Monte Carlo Study

The results of the Monte Carlo study are presented in this section. This study is styled after the study of Paarsch [1983]. The basic model of this study is:

\[ y_j = a + bx_{ij}^0 + u_j. \]

The following items are studied:

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

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

iii) The effect of grouping. Two, five, ten, and twenty are considered as the number of cells.

The values of \( x_{ij}^0 \) are chosen from \([0,20]\), spacing them at equidistant points from 0 to 20 where the distance between the points is determined by the sample size. The value of \( a \) and \( b \) are -10.0 and 1.0 for all cases. 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 OLS, LAD, and \( \hat{\beta} \) where the numbers of cells are two, five, ten and twenty, are studied. The size of each cell is chosen to be equal and is determined by the number of cells.

The findings of the study are:

i) The OLS estimator performs extremely poorly when the distribution is Cauchy.
ii) In all but the 50-sample and 20-cell (2 or 3 observations per cell) cases, the LAD estimator and \( \hat{\beta} \) perform reasonably well over all distributions.


### TABLE I

**SAMPLE SIZE=50, CAUCHY**

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-156.5</td>
<td>1517.8</td>
<td>-28.37</td>
<td>-11.03</td>
<td>6.79</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>11.42</td>
<td>101.9</td>
<td>-0.18</td>
<td>1.13</td>
<td>3.01</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.57</td>
<td>5.43</td>
<td>-13.84</td>
<td>-9.96</td>
<td>-5.41</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.98</td>
<td>0.49</td>
<td>0.65</td>
<td>1.03</td>
<td>1.33</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.61</td>
<td>5.75</td>
<td>-13.46</td>
<td>-10.18</td>
<td>-6.35</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.98</td>
<td>0.52</td>
<td>0.66</td>
<td>1.03</td>
<td>1.29</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.30</td>
<td>6.04</td>
<td>-12.73</td>
<td>-10.02</td>
<td>-5.43</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.95</td>
<td>0.55</td>
<td>0.65</td>
<td>0.96</td>
<td>1.30</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.63</td>
<td>8.90</td>
<td>-14.67</td>
<td>-10.32</td>
<td>-5.23</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.06</td>
<td>0.77</td>
<td>0.60</td>
<td>0.98</td>
<td>1.41</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-4.98</td>
<td>66.60</td>
<td>-20.99</td>
<td>-10.19</td>
<td>0.73</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.65</td>
<td>5.34</td>
<td>0.16</td>
<td>0.94</td>
<td>1.41</td>
</tr>
</tbody>
</table>

### TABLE II

**SAMPLE SIZE=50, LAPLACE**

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.78</td>
<td>4.26</td>
<td>-12.80</td>
<td>-9.67</td>
<td>-6.84</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.38</td>
<td>0.70</td>
<td>1.02</td>
<td>1.28</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.72</td>
<td>3.58</td>
<td>-12.49</td>
<td>-9.97</td>
<td>-6.95</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.32</td>
<td>0.77</td>
<td>1.02</td>
<td>1.25</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.74</td>
<td>3.78</td>
<td>-12.26</td>
<td>-10.10</td>
<td>-7.62</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.34</td>
<td>0.78</td>
<td>1.02</td>
<td>1.19</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.55</td>
<td>3.79</td>
<td>-11.74</td>
<td>-9.95</td>
<td>-7.00</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.97</td>
<td>0.33</td>
<td>0.77</td>
<td>0.97</td>
<td>1.20</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.36</td>
<td>0.74</td>
<td>0.99</td>
<td>1.26</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.94</td>
<td>4.30</td>
<td>-12.94</td>
<td>-10.25</td>
<td>-7.01</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.38</td>
<td>0.70</td>
<td>1.01</td>
<td>1.30</td>
</tr>
</tbody>
</table>
### TABLE III
SAMPLE SIZE=50, NORMAL

<table>
<thead>
<tr>
<th># of cells</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.76</td>
<td>3.20</td>
<td>-12.27</td>
<td>-9.65</td>
<td>-7.29</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.28</td>
<td>0.74</td>
<td>0.99</td>
<td>1.19</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.70</td>
<td>3.96</td>
<td>-13.02</td>
<td>-9.97</td>
<td>-6.64</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.34</td>
<td>0.73</td>
<td>1.02</td>
<td>1.27</td>
</tr>
<tr>
<td>( \hat{\beta} ) 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.76</td>
<td>4.18</td>
<td>-12.70</td>
<td>-10.22</td>
<td>-7.10</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.37</td>
<td>0.74</td>
<td>1.02</td>
<td>1.23</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.54</td>
<td>3.87</td>
<td>-11.74</td>
<td>-9.91</td>
<td>-6.84</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.97</td>
<td>0.33</td>
<td>0.76</td>
<td>0.97</td>
<td>1.22</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.94</td>
<td>3.93</td>
<td>-12.45</td>
<td>-10.30</td>
<td>-7.22</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.36</td>
<td>0.73</td>
<td>0.98</td>
<td>1.25</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.92</td>
<td>3.57</td>
<td>-12.47</td>
<td>-10.06</td>
<td>-7.71</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.31</td>
<td>0.73</td>
<td>1.01</td>
<td>1.23</td>
</tr>
</tbody>
</table>

### TABLE IV
SAMPLE SIZE=100, CAUCHY

<table>
<thead>
<tr>
<th># of cells</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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-29.23</td>
<td>249.7</td>
<td>-28.65</td>
<td>-11.15</td>
<td>4.98</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.88</td>
<td>27.08</td>
<td>-0.40</td>
<td>1.28</td>
<td>2.76</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.05</td>
<td>3.72</td>
<td>-12.59</td>
<td>-9.63</td>
<td>-8.01</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.32</td>
<td>0.81</td>
<td>0.97</td>
<td>1.20</td>
</tr>
<tr>
<td>( \hat{\beta} ) 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.99</td>
<td>4.03</td>
<td>-13.29</td>
<td>-9.16</td>
<td>-7.65</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.36</td>
<td>0.79</td>
<td>0.98</td>
<td>1.24</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.22</td>
<td>4.05</td>
<td>-12.95</td>
<td>-10.02</td>
<td>-7.76</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.37</td>
<td>0.78</td>
<td>1.01</td>
<td>1.28</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.81</td>
<td>4.12</td>
<td>-12.81</td>
<td>-9.96</td>
<td>-7.68</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.98</td>
<td>0.37</td>
<td>0.70</td>
<td>0.98</td>
<td>1.17</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.98</td>
<td>0.44</td>
<td>0.72</td>
<td>0.95</td>
<td>1.27</td>
</tr>
<tr>
<td># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.08</td>
<td>3.10</td>
<td>-12.43</td>
<td>-10.25</td>
<td>-7.95</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>2.28</td>
<td>0.81</td>
<td>0.99</td>
<td>1.19</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.04</td>
<td>2.45</td>
<td>-11.67</td>
<td>-9.77</td>
<td>-8.74</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.21</td>
<td>0.88</td>
<td>0.98</td>
<td>1.13</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.24</td>
<td>0.86</td>
<td>0.99</td>
<td>1.16</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.14</td>
<td>2.67</td>
<td>-11.93</td>
<td>-10.02</td>
<td>-8.54</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.24</td>
<td>0.86</td>
<td>1.01</td>
<td>1.18</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.92</td>
<td>2.54</td>
<td>-11.86</td>
<td>-9.97</td>
<td>-8.57</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.23</td>
<td>0.83</td>
<td>0.99</td>
<td>1.10</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.98</td>
<td>2.96</td>
<td>-12.64</td>
<td>-9.53</td>
<td>-8.05</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>0.99</td>
<td>0.25</td>
<td>0.82</td>
<td>0.99</td>
<td>1.17</td>
</tr>
</tbody>
</table>

**TABLE VI**

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.03</td>
<td>2.31</td>
<td>-11.90</td>
<td>-10.15</td>
<td>-8.53</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.21</td>
<td>0.85</td>
<td>0.97</td>
<td>1.15</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.04</td>
<td>2.84</td>
<td>-12.03</td>
<td>-9.70</td>
<td>-8.42</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.24</td>
<td>0.85</td>
<td>0.97</td>
<td>1.16</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.99</td>
<td>3.08</td>
<td>-12.56</td>
<td>-9.36</td>
<td>-8.12</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.28</td>
<td>0.85</td>
<td>0.99</td>
<td>1.18</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.15</td>
<td>2.94</td>
<td>-12.30</td>
<td>-9.99</td>
<td>-8.27</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.26</td>
<td>0.83</td>
<td>1.01</td>
<td>1.21</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.95</td>
<td>2.62</td>
<td>-12.07</td>
<td>-9.95</td>
<td>-8.39</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.23</td>
<td>0.84</td>
<td>0.99</td>
<td>1.13</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.99</td>
<td>2.86</td>
<td>-12.49</td>
<td>-9.76</td>
<td>-8.00</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.24</td>
<td>0.82</td>
<td>0.98</td>
<td>1.18</td>
</tr>
</tbody>
</table>
### TABLE VII
SAMPLE SIZE=200, CAUCHY

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-44.45</td>
<td>339.5</td>
<td>-32.52</td>
<td>-13.47</td>
<td>0.06</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>2.87</td>
<td>23.02</td>
<td>-0.48</td>
<td>1.31</td>
<td>2.67</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.06</td>
<td>2.83</td>
<td>-11.90</td>
<td>-10.15</td>
<td>-8.71</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.20</td>
<td>0.86</td>
<td>1.03</td>
<td>1.16</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.03</td>
<td>0.26</td>
<td>0.85</td>
<td>1.02</td>
<td>1.19</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.27</td>
<td>2.60</td>
<td>-11.88</td>
<td>-10.61</td>
<td>-8.32</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.03</td>
<td>0.23</td>
<td>0.83</td>
<td>1.48</td>
<td>1.19</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.17</td>
<td>2.42</td>
<td>-11.81</td>
<td>-10.32</td>
<td>-8.56</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.03</td>
<td>0.21</td>
<td>0.87</td>
<td>1.05</td>
<td>1.16</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-9.98</td>
<td>2.87</td>
<td>-12.41</td>
<td>-9.64</td>
<td>-7.95</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.25</td>
<td>0.85</td>
<td>0.97</td>
<td>1.17</td>
</tr>
</tbody>
</table>

### TABLE VIII
SAMPLE SIZE=200, LAPLACE

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.09</td>
<td>1.94</td>
<td>-11.66</td>
<td>-9.83</td>
<td>-8.92</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.18</td>
<td>0.89</td>
<td>0.99</td>
<td>1.13</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.05</td>
<td>1.52</td>
<td>-11.20</td>
<td>-10.10</td>
<td>-9.18</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.13</td>
<td>0.91</td>
<td>1.02</td>
<td>1.11</td>
</tr>
<tr>
<td>( \hat{\beta} )</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.13</td>
<td>1.84</td>
<td>-11.16</td>
<td>-10.24</td>
<td>-9.10</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.17</td>
<td>0.91</td>
<td>1.01</td>
<td>1.13</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.18</td>
<td>1.71</td>
<td>-11.20</td>
<td>-10.39</td>
<td>-8.91</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.15</td>
<td>0.89</td>
<td>1.03</td>
<td>1.12</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.11</td>
<td>1.60</td>
<td>-11.19</td>
<td>-10.22</td>
<td>-8.92</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.14</td>
<td>0.92</td>
<td>1.03</td>
<td>1.11</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-10.0</td>
<td>-10.01</td>
<td>1.78</td>
<td>-11.58</td>
<td>-9.78</td>
<td>-8.80</td>
</tr>
<tr>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.16</td>
<td>0.90</td>
<td>0.99</td>
<td>1.11</td>
</tr>
</tbody>
</table>
### TABLE IX
SAMPLE SIZE=200, NORMAL

<table>
<thead>
<tr>
<th># of cells</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>OLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.06</td>
<td>1.48</td>
<td>-11.23</td>
<td>-9.82</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.00</td>
<td>0.14</td>
<td>0.91</td>
<td>0.99</td>
</tr>
<tr>
<td>LAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.06</td>
<td>1.84</td>
<td>-11.51</td>
<td>-10.13</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.16</td>
<td>0.89</td>
<td>1.02</td>
</tr>
<tr>
<td>( \hat{\beta} ) 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.16</td>
<td>2.21</td>
<td>-11.45</td>
<td>-10.30</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.03</td>
<td>0.21</td>
<td>0.88</td>
<td>1.02</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.21</td>
<td>1.97</td>
<td>-11.49</td>
<td>-10.48</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.03</td>
<td>0.17</td>
<td>0.87</td>
<td>1.04</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.13</td>
<td>1.77</td>
<td>-11.39</td>
<td>-10.19</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.02</td>
<td>0.16</td>
<td>0.89</td>
<td>1.04</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>-10.0</td>
<td>-10.03</td>
<td>1.83</td>
<td>-11.56</td>
<td>-9.84</td>
</tr>
<tr>
<td></td>
<td>b</td>
<td>1.0</td>
<td>1.01</td>
<td>0.16</td>
<td>0.90</td>
<td>1.00</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 as follows:

1. \( L \): the number of cells (the first-stage estimator).
2. \( L^* \): the number of cells (the second-stage estimator).
3. \( K \): the number of independent variables (excluding the constant term).
4. \( N \): the number of observations.
5. \( n_i \): the number of observations in the \( i \)-th cell.
6. \( X_{i}^{O} \): the value of independent variables representing the \( i \)-th cell (excluding the constant term).
7. \( X_{i}^{O} \) = \( (1, X_{i}^{O}) \).
8. \( y_j \): the dependent variable.
9. \( X_{j}^{O} \): the independent variables.
10. \( X_{j}^{O} \) = \( (1, X_{j}^{O}) \).
11. \( \bar{X}_{1}^{O} = \frac{1}{L} \sum_{k=1}^{L} X_{k}^{O} \).
12. \( I_i \): the index set of the \( i \)-th cell (if \( j \in I_i \), the \( j \)-th observation belongs to the \( i \)-th cell).
13. \( x_{1}^{*} = E \{ x_{j}^{O} | j \in I_i \} \).
14. \( X_{1}^{*} = (1, x_{1}^{*}) \).
15. \( y_j(a) = y_j - (x_{j}^{O} - \bar{X}_{1}^{O})'a \), where \( j \) is the cell to which the \( j \)-th observation belongs and \( a \) is a \( K \times 1 \) vector.
16. \( A_{1}(a) = \{ y_j(a) | j \in I_i \} \).
17. \( M[A] \): the sample estimator of median of the set \( A \).
18. \( \bar{X}^{O} = (\bar{X}_{1}^{O}, \bar{X}_{2}^{O}, \ldots, \bar{X}_{L}^{O}) \).
19. \( X' = (X_1, X_2, \ldots, X_{L^*}) \).

20. \( \mu_i(\alpha) = \begin{cases} M[A_i(\alpha)] & \text{if the } i\text{-th cell is nonempty} \\ 0 & \text{otherwise.} \end{cases} \)

21. \( \mu(\alpha)' = [\mu_1(\alpha), \mu_2(\alpha), \ldots, \mu_{\Lambda}(\alpha)] \).
where \( \Lambda = L \) (the first stage), \( L^* \) (the second stage).

22. \( \theta_1 = M\{u_j | j \in I_1\} \), where \( u_j \) is the error term of the regression equation.

23. \( \theta' = (\theta_1, \theta_2, \ldots, \theta_{\Lambda}) \).

24. \( T(\alpha) = (\bar{\chi}^0, \bar{\chi}^0)^{-1} \bar{\chi}^0 \mu(\alpha) \).

25. \( \beta_1 \): the first-stage estimator which is either a fixed point of \( T(\alpha) \) or zero.

26. \( \Sigma = \begin{bmatrix} n_1/N & 0 \\ n_2/N & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_{L^*}/N \end{bmatrix} \).

27. \( \hat{\beta} \): the second stage estimator.
\( \hat{\beta} = (X' \Sigma X)^{-1} X' \Sigma \mu(\beta_1) \).

28. \( \rho(\alpha, \alpha^*) = \sum_{k=1}^{K} |\alpha_k - \alpha_k^*| \), where \( \alpha \) and \( \alpha^* \) are the K-th dimensional vectors and \( \alpha_k \) and \( \alpha_k^* \) are the k-th elements.
Appendix B

If $u_j$ satisfy the following condition, we can apply the proposed method and obtain a consistent estimator of $\beta$. The definition of an estimator and the proof of consistency are exactly the same as in the median case. The conditions are:

i) $m(F_j) = 0$ for all $j$ where $F_j$ is the distribution function of $u_j$ and $m$ is the operator in question. (For example, $m(F)$ denotes the population median, a% quantile, etc.)

ii) For any sequence of real numbers $\{a_j\} = \{a_1, a_2, \ldots, a_N\}$, there exists a sample estimator $M$ of $m$ which satisfies:

\[
M\{a_j + c\} = M\{a_j\} + c,
\]

\[
M\{a_j + e_j\} > M\{a_j\} \text{ if } e_j > 0 \text{ for all } j,
\]

\[
M\{a_j + e_j\} < M\{a_j\} \text{ if } e_j < 0 \text{ for all } j,
\]

\[
M\{a_j + e_j\} < M\{a_j\} \text{ if } e_j < 0 \text{ for all } j, \text{ and}
\]

where $\{e_j\}$ is a sequence of real numbers.

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

The median, a% quantile and $L_p$ ($p>1$) estimators satisfy these conditions. However, the mode does not satisfy these conditions.

Therefore, this paper includes models such as

\[
\text{Median}(y_j) = x_j ' \beta,
\]

a% quantile$(y_j) = x_j ' \beta$, and

\[
E(y_j) = x_j ' \beta.
\]

But it does not include,

\[
\text{Mode}(y_j) = x_j ' \beta.
\]
Appendix C

[Proof of Theorem 4.2]

First, the special case is proved. Let the density of $u_j$ be

\[(C.1) \quad f(u) = \begin{cases} 1 & \text{if } -1/2 < u < 1/2 \\ 0 & \text{otherwise} \end{cases}\]

Let $\xi_j = \sqrt{\tau} u_j$. Then the density of $\xi_j$ is

\[(C.2) \quad f_{\tau}(\xi) = \begin{cases} 1/\sqrt{\tau} & \text{if } -\sqrt{\tau} < \xi < \sqrt{\tau}/2 \\ 0 & \text{otherwise} \end{cases}\]

Let $G(v)$ be the distribution function of $v_j$ and $V$ be the support of $G$. $V$ is a bounded subset of $R^K$. From Theorem 4.1, $\sqrt{\tau} \hat{\theta}_\tau$ converges to a normal variable with mean zero and finite variance. Therefore, for any $\varepsilon > 0$, there exist $M_\varepsilon$ and $\tau_0$ such that

\[(C.3) \quad P\left[ \sup_{v \in V} |\sqrt{\tau} v' \mathbf{v}_\tau | > M_\varepsilon /4 \right] < \varepsilon /4,\]

\[P[|\sqrt{\tau} \hat{\theta}_\tau | > M_\varepsilon /2] < \varepsilon /4,\]

\[M_\varepsilon > 4\varepsilon, \text{ and} \]

\[\sqrt{\tau} > 10M_\varepsilon \quad \text{for } \tau > \tau_0.\]

Define the space $\Gamma$ as

\[(C.4) \quad \Gamma = \{(b, M^*) | \sup_{v \in V} |v'b| < M_\varepsilon /2, |M^*| < M_\varepsilon \} .\]

Let
(C.5) \[ z_j(b,M^*) = \begin{cases} 1 & \text{if } \sqrt{\tau} u_j < M^*, \sqrt{\tau} u_j + b'v_j > M^* \\ -1 & \text{if } \sqrt{\tau} u_j > M^*, \sqrt{\tau} u_j + b'v_j < M^* \\ 0 & \text{otherwise.} \end{cases} \]

Since \( u_j \) and \( v_j \) are independent for all \( j \), \( (u_j,v_j) \) and \( (u_k,v_k) \) are independent for all \( j \neq k \), and \( \mathbb{E} v_j = 0 \), for any \( (b,M^*) \in \Gamma \),

(C.6) \[ \mathbb{E} z_j(b,M^*) = \frac{1}{\sqrt{\tau}} \int_{M^*-b'v}^{M^*} f_\tau(\xi) \, d\xi \, dG(v) \]

\[ = \int_{\mathbb{R}} b'v/\sqrt{\tau} \, dG(v) \]

\[ = 0. \]

(C.7) \[ \text{Var } z_j(b,M^*) = \frac{1}{\sqrt{\tau}} \int_{M^*-|b'v|}^{M^*} f_\tau(\xi) \, d\xi \, dG(v) \]

\[ = \int_{\mathbb{R}} |b'v|/\sqrt{\tau} \, dG(v) \]

\[ = B(b)/\sqrt{\tau}. \]

By Chebyshev's inequality,

(C.8) \[ \mathbb{P}\left[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} z_j(b,M^*) > \lambda \right] \leq \frac{1}{\lambda^2} \text{Var}\left[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} z_j(b,M^*) \right] \]

\[ = \frac{1}{\lambda^2 \tau} \sum_{j=1}^{\tau} \text{Var } z_j(b,M^*). \]

Let \( B^* = \sup_{b \in \Gamma} B(b) \).

Since \( V \) and \( \Gamma \) are bounded, \( B^* < \infty \).

Clearly for all \( (b,M^*) \in \Gamma \),
(C.9) \[ P[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} z_j(b, M^*) | \lambda ] < \frac{B^*}{\lambda^2 \tau} . \]

Hence for any \( \lambda > 0 \),

\[ \sup_{(b, M^*) \in \Gamma} P[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} z_j(b, M^*) | \lambda ] \rightarrow 0 \text{ as } \tau \rightarrow \infty . \]

Let

\[ \eta(M^*, \zeta) = \begin{cases} 1 & \text{if } \sqrt{\tau} u_j \in (M^* - \zeta, M^*) \\ 0 & \text{otherwise.} \end{cases} \]

If \( |\zeta| < M^*/2 \) and \( |M^*| < M^* \),

\[ E \eta(M^*, \zeta) = \frac{\zeta}{\sqrt{\tau}} , \]

and

\[ \text{Var } \eta_j(M^*, \zeta) = E[ \eta_j(M^*, \zeta) - \frac{\zeta}{\sqrt{\tau}} ]^2 \]

\[ = \int_{M^*-\zeta}^{M^*} (1 - \zeta/\sqrt{\tau})^2 / \sqrt{\tau} \, du + \int_{\zeta/\sqrt{\tau}}^{\sqrt{\tau}/2} (\zeta/\sqrt{\tau})^2 / \sqrt{\tau} \, du \]

\[ + \int_{-\sqrt{\tau}/2}^{\zeta/\sqrt{\tau}} (\zeta/\sqrt{\tau})^2 / \sqrt{\tau} \, du \]

\[ = \frac{\zeta}{\sqrt{\tau}} - \zeta^2 / \tau . \]

By Chebyshev's inequality,

\[ P[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} \eta_j(M^*, \zeta) - \frac{\zeta}{\sqrt{\tau}} ] < \frac{1}{\lambda^2} \text{Var } \eta_j(M^*, \zeta) \]

\[ = \frac{1}{\lambda^2} \frac{\zeta}{\sqrt{\tau}} - \zeta^2 / \tau \]

\[ = \frac{(\zeta/\sqrt{\tau} - \zeta^2 / \tau)}{\lambda^2} . \]

Since \( \zeta/\sqrt{\tau} - \zeta^2 / \tau \) does not depend on \( M^* \), for any \( \lambda > 0 \),
(C.15) \[ \sup_{M^* \in \Gamma} \mathbb{P}\left[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} \eta_j(M^*, \zeta) - \zeta \mid \lambda \right] \to 0. \]

It is shown that \[ \sum_{j=1}^{\tau} z_j(b, M^*)/\sqrt{\tau} \] and \[ \sum_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau} \] converge to 0 and \( \zeta \) in probability uniformly in \( \Gamma \).

First it is proved that

\[ \mathbb{P}\left[ \sup_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau - \zeta} \mid \lambda \right] \to 0 \text{ for any } \lambda > 0 \text{ as } \tau \to \infty. \]

The proof is a modification of Amemiya's proof [1985, p.116].

Partition \( \Gamma \) into nonoverlapping regions \( \Gamma_1, \Gamma_2, \ldots, \Gamma_n \) in such a way that the distance between any two points within each \( \Gamma_i \) goes to zero as \( n \) goes to infinity. Let \( (b_i, M_i^*) \) be an arbitrary sequence such that \( (b_i, M_i^*) \in \Gamma \), \( i = 1, 2, \ldots, n \). Then for any \( \lambda > 0 \),

(C.16) \[ \mathbb{P}\left[ \sup_{M^* \in \Gamma} \left| \sum_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau - \zeta} \right| > \lambda \right] \]

\[ < \mathbb{P}\left[ \bigcup_{i=1}^{n} \sup_{M^* \in \Gamma_i} \left| \sum_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau - \zeta} \right| > \lambda \right] \]

\[ < \sum_{i=1}^{n} \mathbb{P}\left[ \sup_{M^* \in \Gamma_i} \left| \sum_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau - \zeta} \right| > \lambda \right] \]

\[ < \sum_{i=1}^{n} \mathbb{P}\left[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} \eta_j(M^*, \zeta) - \zeta \mid \lambda/2 \right] \]

\[ + \sum_{i=1}^{n} \mathbb{P}\left[ \frac{1}{\sqrt{\tau}} \sum_{j=1}^{\tau} \eta_j(M^*, \zeta) - \zeta \mid \lambda/2 \right]. \]
Let
\[ \phi_j^i = \sup_{M^* \in \Gamma_i} |\eta_j(M^*, \zeta) - \eta_j(M^*_i, \zeta)|, \]
\[ \overline{M}^* = \sup_{M^* \in \Gamma_i} M^*, \quad \underline{M}^* = \inf_{M^* \in \Gamma_i} M^*, \quad \text{and} \quad m_i = \overline{M}^* - \underline{M}^*. \]

Then
\[ \phi_j^i = \begin{cases} 1 & \text{if } \eta_j(M^*_i, \zeta) \neq \sup_{M^* \in \Gamma_i} \eta_j(M^*, \zeta) \quad \text{or} \quad \eta_j(M^*_i, \zeta) \neq \inf_{M^* \in \Gamma_i} \eta_j(M^*, \zeta) \\ 0 & \text{otherwise.} \end{cases} \]

Make \( n \) large enough so that \( m_i < \epsilon / 2 \).

Then
\[ (C.17) \quad \frac{\tau}{\sqrt{\pi}} \sum_{j=1}^{i \phi_j^i} \text{ number of } \sqrt{\tau} u_j \text{'s in } [\overline{M}^*_i, M^*_i] \text{ and } [\underline{M}^*_i - \zeta, M^*_i - \zeta]. \]

Let
\[ (C.18) \quad \Phi_i = \frac{1}{\sqrt{\pi}} \text{ (number of } \sqrt{\tau} u_j \text{'s in } [\overline{M}^*_i, M^*_i] \text{ and } [\underline{M}^*_i - \zeta, M^*_i - \zeta]). \]

Then
\[ (C.19) \quad E \Phi_i = m_i, \quad \text{and} \]
\[ (C.20) \quad \text{Var} \Phi_i = m_i / \sqrt{\pi - 4m_i^2 / \tau}. \]

Hence by Chebyshev's inequality,
\[ (C.21) \quad P[|\Phi_i - 2m_i| > \lambda] < \text{Var} \Phi_i / \lambda^2 \]
\[ = \left( \frac{2m_i}{\sqrt{\pi - 4m_i^2 / \tau}} \right) / \lambda^2. \]

Since \( \Phi_i > 0 \),
\[ P[\Phi_i > 2m_i + \lambda / 4] < (4/\lambda)^2 \left( \frac{2m_i}{\sqrt{\pi - 4m_i^2 / \tau}} \right). \]

Therefore,
\[ (C.22) \quad \sum_{i=1}^{n} P[\Phi_i > 2m_i + \lambda / 4] < \left( \frac{4}{\lambda} \right)^2 \sum_{i=1}^{n} \left( \frac{2m_i}{\sqrt{\pi - 4m_i^2 / \tau}} \right). \]
Since \( \phi_i = \text{constant} \) and \( \phi_i^2 \leq \infty \), if we make \( \text{Max } m_i < \lambda/8 \), there exist \( \tau_1 \) and \( H \leq \infty \) such that

\[
\sum_{i=1}^{n} P[\phi_i > \lambda] < H/\sqrt{\tau} \quad \text{for } \tau > \tau_1.
\]

Since \( \sum_{j=1}^{\tau} \sup_{M^* \in \Gamma_i} |\eta_j(M^*, \zeta) - \eta_j(M_i^*, \zeta)|/\sqrt{\tau} = \sum_{j=1}^{\tau} \phi_j^2/\sqrt{\tau} < \phi_i \),

\[
P[\phi_i > \lambda/2] \geq P\left[ \sum_{j=1}^{\tau} \sup_{M^* \in \Gamma_i} |\eta_j(M^*, \zeta) - \eta_j(M_i^*, \zeta)|/\sqrt{\tau} > \lambda/2 \right].
\]

Therefore,

\[
\sum_{i=1}^{n} P[\sum_{j=1}^{\tau} \sup_{M^* \in \Gamma_i} |\eta_j(M_i^*, \zeta)|/\sqrt{\tau} > \lambda/2] < \lambda/2 \quad \text{for } \tau > \tau_1
\]

and from (A.15),

\[
\sum_{i=1}^{n} P[\sum_{j=1}^{\tau} \eta_j(M_i^*, \zeta)/\sqrt{\tau} > \lambda/2] \to 0 \quad \text{as } \tau \to \infty.
\]

Hence there exists \( \tau_2 \) such that,

\[
\sum_{i=1}^{n} P[\sum_{j=1}^{\tau} \eta_j(M_i^*, \zeta)/\sqrt{\tau} > \lambda/2] < \lambda/2 \quad \text{for } \tau > \tau_2.
\]

From (A.16), (A.25), and (A.27) for any \( \lambda > 0 \), there exists \( \tau_3 \) such that

\[
P[\sup_{M^* \in \Gamma} \sum_{j=1}^{\tau} \eta_j(M^*, \zeta)/\sqrt{\tau} > \lambda] < \lambda \quad \text{for } \tau > \tau_3.
\]

Hence,
Next it is shown that $P \sup_{(b, M^*) \in \Gamma} \frac{\tau}{j=1} \sum_{j=1}^{\tau} z_j(b, M^*)/|\tau| >\lambda \to 0$ as $\tau \to \infty$. As before,

\[
(C.29) \quad P \sup_{(b, M^*) \in \Gamma} \frac{\tau}{j=1} \sum_{j=1}^{\tau} z_j(b, M^*)/|\tau| >\lambda < \sum_{i=1}^{n} P \sup_{i=1}^{\tau} z_j(b_i, M^*)/|\tau| >\lambda/2 | + \sum_{i=1}^{n} P \sup_{i=1}^{\tau} z_j(b_i, M^*)/|\tau| >\lambda/2 |.
\]

Partition $V$ into $k$ nonoverlapping regions $V_1, V_2, \ldots, V_k$ in such a way that the distance between any two points within each $V_i$ goes to zero as $k \to \infty$. Let

\[
(C.30) \quad \phi_{ijk} = \begin{cases} 1 & \text{if } v_j \in V_k \text{ and } \sup_{(b, M^*) \in \Gamma_i} z_j(b, M^*) \neq z_j(b_i, M^*), \\
\inf_{(b, M^*) \in \Gamma_i} z_j(b, M^*) \neq z_j(b_i, M^*), & \text{otherwise}.
\end{cases}
\]

Here,

\[
(C.31) \quad \frac{\tau}{j=1} \phi_{ijk} \leq |\text{number of } \sqrt{\tau} u_j \text{ which satisfy (i) } v \in V_j, \\
(ii) \sqrt{\tau} u_j < M, (iii) \sqrt{\tau} u_j + b_k > M, (iv) \sqrt{\tau} u_j + b_k < M | + |\text{number of } \sqrt{\tau} u_j \text{ which satisfy (i) } v_j \in V_k, \\
(ii) \sqrt{\tau} u_j > M, (iii) \sqrt{\tau} u_j + b_k < M, (iv) \sqrt{\tau} u_j + b_k > M |
\]
\[\{\text{number of } \sqrt{\tau} u_j \text{ which satisfy (i) } v_j \epsilon V_k, \]
\[\text{(ii) } \sqrt{\tau} u_j \epsilon [M^*_b, M^*_b] U [M^*_{-1}, M^*_{-1}] U [M^*_1, M^*_1] U [M^*_1, M^*_1] \} .\]

Let

(C.32) \[\Phi_{ik} = \frac{1}{\sqrt{\tau}} \{\text{number of } \sqrt{\tau} u_j \text{ which satisfy (i) } v_j \epsilon V_k, \]
\[\text{(ii) } \sqrt{\tau} u_j \epsilon [M^*_b, M^*_b] U [M^*_1, M^*_1] U [M^*_1, M^*_1] \} .\]

Then

(C.33) \[E \Phi_{ik} = (3m_{-1} + 2 b_{ik}^*) P[V_k], \]
and

(C.34) \[\text{Var } \Phi_{ik} = \{((3m_{-1} + 2b_{ik}^*)/\sqrt{\tau}) - ((3m_{-1} + 2b_{ik}^*)^2/\tau)\} P[V_k]. \]

Let \( \Phi_i = \sum_{k=1}^{\kappa} \Phi_{ik} \).

Since \( \Phi_{ik} \)'s are independent,

(C.35) \[E \Phi_i = \sum_{k=1}^{\kappa} (3m_{-1} + 2b_{ik}^*) P[V_k], \]
and

(C.36) \[\text{Var } \Phi_i = \sum_{i=1}^{\kappa} \{((3m_{-1} + 2b_{ik}^*)/\sqrt{\tau}) - ((3m_{-1} + 2b_{ik}^*)^2/\tau)\} P[V_k]. \]

Since \( m_{-1}, b_{ik}^* \to 0 \) as \( n \) and \( \kappa \to \infty \), the argument of the previous proof follows and \( \frac{1}{\sqrt{\tau}} z_j(b, M^*)/\sqrt{\tau} \) converges to zero in probability uniformly in \( \Gamma \).

Let

(C.37) \[Z(b, M^*) = \sum_{j=1}^{\tau} z_j(b, M^*)/\sqrt{\tau} \]
=the net number of elements which cross \( \zeta = M^* \) from the left, and

\[
(C.38) \quad n(M^*, \varepsilon) = \sum_{j=1}^{\tau} \eta_j(M^*, \varepsilon) / \sqrt{\tau}
\]

=the number of elements in \((M^*-\varepsilon, M^*)\).

Suppose

\[
(C.39) \quad \sqrt{\tau} \left( \hat{\theta}_\tau - \hat{\theta}_\tau + \varepsilon \right) > \varepsilon .
\]

Then

\[
(C.40) \quad Z(\sqrt{\tau} w_\tau, \sqrt{\tau} \hat{\theta}_\tau + \varepsilon) > n(\sqrt{\tau} \hat{\theta}_\tau, \varepsilon) - 1 .
\]

Since \( Z(b, M^*) / \sqrt{\tau} \to 0 \) and \( n(M^*, \varepsilon) \to \varepsilon \) in probability uniformly in \( \Gamma \), there exists \( \tau_1 \) such that,

\[
(C.41) \quad P[ \sup_{(b,M^*) \in \Gamma} Z(b, M^*) / \sqrt{\tau} > \varepsilon / 4] < \varepsilon / 8, \text{ and } n(M^*, \varepsilon) / \sqrt{\tau} < \varepsilon / 8 \text{ for } \tau > \tau_1 .
\]

Therefore,

\[
(C.43) \quad P[ \sup_{(b,M^*) \in \Gamma} Z(b, M^*) / \sqrt{\tau} > \inf_{M^* \in \Gamma} n(M^*, \varepsilon)] < \varepsilon / 4 ,
\]

Suppose \((\sqrt{\tau} w_\tau, \psi) \in \Gamma\) for any \( \psi [\sqrt{\tau} \hat{\theta}_\tau - \varepsilon, \sqrt{\tau} \hat{\theta}_\tau + \varepsilon] \). Then

\[
(C.44) \quad \sup_{(b,M^*) \in \Gamma} Z(b, M^*) > Z(\sqrt{\tau} w_\tau, \sqrt{\tau} \hat{\theta}_\tau + \varepsilon), \text{ and } n(M^*, \varepsilon) / \sqrt{\tau} < \varepsilon / 8 \text{ for } \tau > \tau_1 .
\]

\[
(C.45) \quad \inf_{M^* \in \Gamma} n(M^*, \varepsilon) < n(\sqrt{\tau} \hat{\theta}_\tau + \varepsilon, \varepsilon) .
\]

From (A.44) and (A.45),

\[
(C.46) \quad P[ \sup_{(b,M^*) \in \Gamma} Z(b, M^*) > \inf_{M^* \in \Gamma} n(M^*, \varepsilon)]
\]
\[ P[\sqrt{\frac{m}{n}} \sqrt{\frac{\hat{\theta}}{n}} + \epsilon] \leq \frac{\epsilon}{4} \]

From (A.43) and (A.46),

\[ P[\sqrt{\frac{m}{n}} \sqrt{\frac{\hat{\theta}}{n}} + \epsilon] > n(\sqrt{\frac{\hat{\theta}}{n}} + \epsilon, \epsilon) \leq \frac{\epsilon}{4} \]

and

\[ P[\sqrt{\frac{m}{n}} \sqrt{\frac{\hat{\theta}}{n}} + \epsilon] > n(\sqrt{\frac{\hat{\theta}}{n}} + \epsilon, \epsilon) - 1] + \]

\[ P[\sqrt{\frac{m}{n}} \sqrt{\frac{\hat{\theta}}{n}} + \epsilon] > n(\sqrt{\frac{\hat{\theta}}{n}} + \epsilon, \epsilon) \text{ as } \tau \to \infty. \] Hence,

\[ P[\sqrt{\frac{m}{n}} (\hat{\theta} - \theta) > \epsilon] < \frac{\epsilon}{4} \]

for \( \tau > \tau_1 \).

In the same way there exists \( \tau_2 \) such that

\[ P[\sqrt{\frac{m}{n}} (\hat{\theta} - \theta) < -\epsilon] < \frac{\epsilon}{4} \]

for \( \tau > \tau_2 \).

From (A.48) and (A.49), for \( \tau > \tau_3 = \text{Max}(\tau_1, \tau_2) \),

\[ P[|\sqrt{\frac{m}{n}} (\hat{\theta} - \theta)| > \epsilon] < \frac{\epsilon}{2} \]

For \( \tau > \tau_0 \),

\[ P[\sup_{u \in \Gamma} |\sqrt{\frac{m}{n}} u | > M \frac{\epsilon}{4}] < \frac{\epsilon}{4}, \]

\[ P[|\sqrt{\frac{m}{n}} \hat{\theta} | > M \frac{\epsilon}{4}] < \frac{\epsilon}{4}, \text{ and} \]

\[ M > \epsilon. \]

Therefore,

\[ P[\Gamma(\sqrt{\frac{m}{n}}, \phi)] > 1 - \frac{\epsilon}{2} \text{ for } \tau > \tau_0 \text{ and } \phi \in \Gamma(\sqrt{\frac{\hat{\theta}}{n}} - \epsilon, \sqrt{\frac{\hat{\theta}}{n}} + \epsilon). \]

From (A.50) and (A.51), for any \( \epsilon > 0 \), there exists \( \tau_4 \) such that

\[ P[|\sqrt{\frac{m}{n}} (\hat{\theta} - \theta)| > \epsilon] < \epsilon \text{ for } \tau > \tau_4. \]

Hence,

\[ \sqrt{\frac{m}{n}} (\hat{\theta} - \theta) \to 0. \]
Now the case is proved where \( f \) is the general form of a function which satisfies the assumptions. Let \( f_\tau \) be the density function of \( \xi_j = \sqrt{\tau} u_j \).

Let \(|\xi| < M_\xi \). Since \( \xi/\sqrt{\tau} \to 0 \) as \( \tau \to \infty \) and \( f'(\xi) \) is bounded in the neighborhood of \( \xi=0 \). So, for large \( \tau \),

\[
(C.54) \quad f_\tau(\tau) = f(\xi/\sqrt{\tau})/\sqrt{\tau} = f(0)/\sqrt{\tau} + \xi f'(\xi^*)/\tau = f(0)/\sqrt{\tau} + O(1/\tau)
\]

where \( \xi^* \) is some value between 0 and \( \xi/\sqrt{\tau} \).

Therefore, all the arguments in the previous case follow in the same way. Hence,

\[
P(\sqrt{\tau}(\hat{\theta}_\tau - \tilde{\theta}_\tau) \to 0 \quad \text{if the assumptions are satisfied.}
\]

QED.
Footnotes

1. Some of the conditions may be replaced by other conditions. For example, let \( \{S_i^*\} \) be a sequence of random cells and \( \{X_i^*\} \) be a sequence of random variables. If these sequences converge to the sequences which satisfy (i) through (viii) as \( N \) goes to infinity with probability one, then the first-stage estimator based on \( \{S_i^*\} \) and \( \{X_i^*\} \) still possesses the same properties.

2. In the case where \( K \) is large and \( N \) is not large enough, it is sometimes difficult to make the number of observations per cell large enough under the conditions (i) through (viii). However there usually exist a sequence of nonempty cells \( \{S_i^*\} \), containing more observations in each cell than \( \{S_i\} \), and \( \{X_i^*\} \) which make \( T^* \) a contraction mapping where \( T^* \) is defined based on \( \{S_i^*\} \) and \( \{X_i^*\} \) in the same way as \( T \). In this case we may use \( T^* \) assuming \( \{S_i^*\} \) and \( \{X_i^*\} \) converge to proper sequences as \( N \) goes to infinity with probability one.

3. Let \( D \) be a diagonal matrix whose \( k \)-th diagonal element is \( w_k \). Then (2.1) can be written as \( y_j = \beta_0 + (D^{-1}x_j^0)'D\beta_1 + u_j \). Since the domains of \( x_j^0 \) are fixed and bounded, \( D \) is a fixed matrix for given \( v \). Therefore, the estimation of \( \beta_1 \) and \( D\beta_1 \) are equivalent in this case.

4. Equalities and inequality involving random variables in this paper
are supposed to hold for every possible realization of random variables. However all results are valid if they hold with probability one.
Reference


Kolmogorov, A.N. and S.V. Fomin (1975), Introductory Real Analysis,
Prentice-Hall, Englewood Cliffs, N.J.


ECONOMETRIC WORKSHOP

Technical Reports in This Series


