2007

Estimation of regression coefficients with unequal probability samples

Yu Y. Wu

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation


https://lib.dr.iastate.edu/rtd/15489

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Estimation of regression coefficients with unequal probability samples

by

Yu Y. Wu

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Wayne A. Fuller, Major Professor
Jean D. Opsomer
Tapabrata Maiti
Song X. Chen
Wolfgang Kliemann

Iowa State University
Ames, Iowa
2007

Copyright © Yu Y. Wu, 2007. All rights reserved.
To my family
# TABLE OF CONTENTS

LIST OF TABLES .................................................. v
LIST OF FIGURES .................................................. vii
ACKNOWLEDGEMENTS .............................................. viii

CHAPTER 1 Introduction ........................................... 1
  1.1 Overview ..................................................... 1
  1.2 Objectives ................................................... 1
  1.3 Modeling Framework .......................................... 3
  1.4 Review of Estimation Approaches ............................. 7
    1.4.1 Ordinary Least Squares Estimator ....................... 7
    1.4.2 Probability Weighted Estimator ......................... 8
    1.4.3 Pfeffermann-Sverchkov Estimator ...................... 10
    1.4.4 Instrumental Variable Estimator ....................... 12
    1.4.5 A Test for Importance of Weights ..................... 14
    1.4.6 Preliminary Testing Procedures ....................... 16

CHAPTER 2 Design Consistent Estimation ......................... 20
  2.1 H Estimator .................................................. 20
    2.1.1 Motivation ............................................... 20
    2.1.2 Central Limit Theorem .................................. 20
  2.2 Pfeffermann-Sverchkov Estimator ............................ 30
CHAPTER 3 Instrumental Variable Estimation

3.1 Introduction

3.2 Instrumental Variables for Weighted Samples

3.3 Central Limit Theorem

3.4 A Test for Endogeneity

CHAPTER 4 Preliminary Testing Procedure

4.1 Pretest Procedure Example

4.2 Test for Importance of Weights

4.3 Instrumental Variable Pretest Procedure

CHAPTER 5 Simulation Design

5.1 Monte Carlo Study Set-up

5.2 Estimators for the Population Parameter

5.2.1 Pfeffermann-Sverchkov Estimator

5.2.2 Instrumental Variable Estimator

5.2.3 Preliminary Testing Procedures

5.2.4 Two-step Preliminary Testing Procedure

5.3 Simulation Results

5.3.1 Case One Study

5.3.2 Case Two Study

5.3.3 Case Three Study

5.4 Concluding Remarks

5.5 A Sample Example

CHAPTER 6 Summary and Conclusion

BIBLIOGRAPHY
LIST OF TABLES

Table 5.1 Monte Carlo Correlations for $\text{corr}_1 = \text{corr}(p_i, x_i)$ and $\text{corr}_2 = \text{corr}(p_i, e_i)$ (10,000 samples) ......... 59

Table 5.2 Monte Carlo Means for $E_1 = E\{p_i x_i\}, E_2 = E\{p_i x_i^2\}$ and $E_3 = E\{p_i e_i\}(10,000$ samples) ........... 60

Table 5.3 Monte Carlo Bias Ratio $\frac{\text{Bias}(\hat{\beta}_0)}{[V(\hat{\beta}_0)]^{1/2}}$ for estimators of $\beta_0$ (10,000 samples) ........... 61

Table 5.4 Monte Carlo Bias Ratio $\frac{\text{Bias}(\hat{\beta}_1)}{[V(\hat{\beta}_1)]^{1/2}}$ for estimators of $\beta_1$ (10,000 samples) ........... 62

Table 5.5 Monte Carlo Mean Squared Error ($\times1000$) for estimators of $\beta_0$ (10,000 samples) ..................... 63

Table 5.6 Monte Carlo Mean Squared Error ($\times1000$) for estimators of $\beta_1$ (10,000 samples) ..................... 64

Table 5.7 Monte Carlo Mean Squared Error Ratio $\frac{\text{MSE}(\hat{\beta}_{PW,0})}{\text{MSE}(\hat{\beta}_{PS,0})}$ for estimators of $\beta_0$ (10,000 samples) ........... 65

Table 5.8 Monte Carlo Mean Squared Error Ratio $\frac{\text{MSE}(\hat{\beta}_{PW,1})}{\text{MSE}(\hat{\beta}_{PS,1})}$ for estimators of $\beta_1$ (10,000 samples) ........... 66

Table 5.9 Monte Carlo Mean Squared Error ($\times1000$) for estimators of $\beta_0$ (10,000 samples) ..................... 67

Table 5.10 Monte Carlo Mean Squared Error ($\times1000$) for estimators of $\beta_1$ (10,000 samples) ..................... 68

Table 5.11 Monte Carlo Probability that $|t_{\hat{\beta}_0}| > t_{.025}$ (10,000 samples) ...... 69
Table 5.12  Monto Carlo Probability that $|t_{\beta_1}| > t_{0.025}$ (10,000 samples) . . . 70
Table 5.13  Monte Carlo Mean Squared Error ($\times 1000$) for estimators of $\beta_0$
(10,000 samples) . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
Table 5.14  Monte Carlo Mean Squared Error ($\times 1000$) for estimators of $\beta_1$
(10,000 samples) . . . . . . . . . . . . . . . . . . . . . . . . . . . 72
Table 5.15  Monto Carlo Probability that $|t_{\beta_0}| > t_{0.025}$ (10,000 samples) . . . 73
Table 5.16  Monto Carlo Probability that $|t_{\beta_1}| > t_{0.025}$ (10,000 samples) . . . 73
Table 5.17  Example Sample Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 82
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Plot of $MSE(\hat{\mu})$ vs. $\mu$</td>
<td>51</td>
</tr>
<tr>
<td>5.1</td>
<td>Plot of $r$ vs. $x$</td>
<td>53</td>
</tr>
<tr>
<td>5.2</td>
<td>Plot of MSEs relative to that of $\hat{\beta}_{PS,0}$</td>
<td>74</td>
</tr>
<tr>
<td>5.3</td>
<td>Plot of MSEs relative to that of $\hat{\beta}_{PS,1}$</td>
<td>75</td>
</tr>
<tr>
<td>5.4</td>
<td>Plot of MSEs relative to that of $\hat{\beta}_{IV0,0}$</td>
<td>76</td>
</tr>
<tr>
<td>5.5</td>
<td>Plot of MSEs relative to that of $\hat{\beta}_{IV0,1}$</td>
<td>77</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

This work was supported by Cooperative Agreement No. 68-3A75-4-122 between the USDA Natural Resources Conservation Service and the Center for Survey Statistics and Methodology at Iowa State University. I would like to thank my advisors Dr. Wayne A. Fuller for guiding me on this study and teaching me new statistical ideas and issues in each meeting. I would also like to thank my committee members, Dr. Tapabrata Maiti, Dr. Jean Opsomer, Dr. Song X. Chen, Dr. Wolfgang Kliemann and Dr. Domenico D’Alessandro. I am very thankful to my fellow survey graduate students for providing feedback and encouragement.
CHAPTER 1 Introduction

1.1 Overview

In a simple random sample, an unbiased estimator of the population regression coefficient is the ordinary least squares (OLS) estimator, and an estimator of its variance is easy to calculate. In many surveys, the elements enter the sample with unequal probabilities. In these cases, the sampling weights, commonly the inverses of the selection probabilities, can be used to construct the probability weighted (PW) estimator. When the weights are related to the values of the response variables, after conditioning on the independent variables in the model, the sampling process becomes informative and the model holding for the sample data is different from the model holding in the population. When the selection probabilities are related to the error terms, the use of the OLS estimator can yield large biases. In complex analyses such as regression, the weighted estimator requires a more complicated calculation, and often gives a larger variance than the unweighted version of the estimator. The OLS estimator and the PW estimator are straightforward procedures, but for complex sampling designs, the OLS estimator and the PW estimator do not always perform well.

1.2 Objectives

One objective is to develop consistent weighted estimators that are more efficient than the PW estimator under complex sampling designs. The alternative estimators are
based on a superpopulation model with error variances determined by values of a co-
variate. Procedures include a design consistent estimator based on estimated variances,
the Pfeffermann-Sverchkov estimator, and an instrumental variable estimator. We will
construct a testing procedure for the importance of weights and discuss an estimation
strategy. If the test statistic is not significant, the unweighted estimator is used. When
the testing procedure indicates that the weighted analysis is preferred, we use a con-
sistent weighted estimator that is more efficient than the PW estimator. Preliminary
testing (pretest) procedures are procedures in which a test of a model assumption is
used to decide between two estimation procedures. We will develop pretest procedures
to obtain a compromise between the unweighted estimator and the weighted estimator.

This thesis is organized as follows. In Section 1.3 the regression models are presented.
In Section 1.4 we briefly review two common estimators, introduce some alternative es-
timation procedures, and describe a test for the importance of weights and a pretest
procedure. In Chapter 2 we discuss two proposed regression estimators in detail. In sec-
tion 2.1 we describe a design consistent estimator based on estimated variances and give
some limiting properties of this design consistent estimator. In section 2.2 we describe
the Pfeffermann-Sverchkov estimator. In Chapter 3 we introduce instrumental variable
estimators, describe some limiting properties, and describe a test for endogeneity for
the instrumental variable procedure. In Chapter 4 pretest procedures are discussed in
detail. We describe a pretest procedure based on the test for importance of weights
and a pretest procedure based on the test for endogeneity in the instrumental variable
procedure. Chapter 5 contains the details of an example based on a simulated data and
a Monte Carlo simulation study designed to illustrate the performance of the alterna-
tive estimators and test statistics, and to compare the alternative estimators with the
OLS and PW estimators described in Section 1.4. The main findings of this study are
discussed in Chapter 6.
1.3 Modeling Framework

Survey data can be viewed as the output of the two random processes: the process generating the values of the finite population from a superpopulation, known as the ‘superpopulation model’, and the process selecting a sample from the finite population, referred to as the ‘sample selection mechanism’. See Pfeffermann et al (1998).

We assume the finite population to be generated by a random process, called the superpopulation. We will use script \( \mathcal{F} \) to denote the finite population, \( U \) to denote the set of indices of the finite population, and \( A \) to denote the set of indices of the sample. We assume that there is a function \( p(\cdot) \) such that \( p(A) \) gives the probability of selecting sample \( A \) from \( U \).

Suppose we have a superpopulation which is used to generate an infinite sequence of \( y \) values, \( y_1, y_2, y_3, \ldots \), where \( y_k = (y_k, x_k) \) is the value tied to the \( k \)-th element. Thus \( \{y_1, y_2, \ldots\} \) is a sequence of iid \((\mu, \Sigma)\) random variables. Let \( \theta \) be a superpopulation parameter. Consider a sequence of populations \( U_1, U_2, U_3, \ldots \), where \( U_N \) consists of \( N \) elements from the infinite sequence of elements, that is, \( U_N = \{1,2,\ldots,N\} \). Let \( \theta_N \) be an estimator of \( \theta \) based on \( U_N \). For each population \( U_N \), a sample \( A_N \) of size \( n_N \) is selected. We assume that \( n_1 < n_2 < n_3 < \ldots \). Thus, as \( N \to \infty \), \( n_N \to \infty \). Let \( \hat{\theta} \) be a finite sample estimator of \( \theta_N \), based on the observed \( y_k \) values, for \( k \in A_N \).

Consider a regression model relating \( y_i \) to \( x_i \) as

\[
y_i = x_i' \beta + e_i, \tag{1.1}
\]

where \( e_i \) are independent \((0, \sigma^2)\) random variables independent of \( x_j \) for all \( i \) and \( j \). The model for the finite population can be written as

\[
y_N = X_N \beta + e_N, \tag{1.2}
\]

\[
e_N \sim (0, I_N \sigma^2),
\]
where $\mathbf{y}_N$ is the $N$ dimensional vector of values for the dependent variable, $\mathbf{X}_N$ is the $N \times k$ matrix of values of the explanatory variables, and the error vector $\mathbf{e}_N$ is the $N$ dimensional vector which is independent of $\mathbf{X}_N$.

Assume a simple random sample (SRS) of size $n$ is selected from the finite population. Then we can write the model for the sample as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e},$$

$$\mathbf{e} \sim (0, \mathbf{I}\sigma^2),$$

where $\mathbf{y}$ is the $n$ dimensional column vector of observations, $\mathbf{X}$ is the $n \times k$ matrix of observations on the explanatory variables, and $\mathbf{e}$ is the $n$ dimensional error vector. Because the sample is a simple random sample, $\mathbf{e}$ is independent of $\mathbf{X}$.

Let $E\{\cdot|\mathcal{F}\}$ be the average overall possible samples under the design for the particular finite population $\mathcal{F}$. The conditional expectation $E\{\cdot|\mathcal{F}\}$ is called the design expectation in survey sampling. Let $E\{\cdot\} = E\{E(\cdot|\mathcal{F})\}$ be the overall average over all possible samples from all possible finite populations. Let $V\{\cdot|\mathcal{F}\}$ be the analogous design variance and let $V\{\cdot\}$ be the analogous overall variance. (Fuller, 2006, pp.17)

Let $E\{\cdot|\mathbf{X}\}$ be the conditional expectation given the explanatory variable $\mathbf{X}$. Let $V\{\cdot|\mathbf{X}\}$ be the analogous conditional variance.

We will use concepts of order as used in real analysis. Let $a_n$ be a sequence of real numbers and $g_n$ be a sequence of positive real numbers.

We say $a_n$ is smaller order than $g_n$ and write

$$a_n = o(g_n)$$

if

$$\lim_{n \to \infty} g_n^{-1} a_n = 0.$$ 

We say $a_n$ is at most of order $g_n$ and write

$$a_n = O(g_n)$$
if there exists a real number $M$ such that

$$g_n^{-1}|a_n| \leq M$$

for all $n$.

For sequences of random variables, we will use the definitions of order in probability. First we will give the concept of convergence in probability.

The sequence of random variables $X_n$ converges in probability to the random variable $X$, and we write

$$p \lim X_n = X,$$

if for every $\varepsilon > 0$

$$\lim_{n \to \infty} P\{|X_n - X| > \varepsilon\} = 0.$$

We use $O_p$ to denote at most of order in probability. Let $X_n$ be a sequence of random variables and $g_n$ be a sequence of positive real numbers. If for every fixed $\varepsilon > 0$, there exists a positive real number $M_\varepsilon$ such that

$$P\{|X_n| \geq M_\varepsilon g_n\} \leq \varepsilon$$

for all $n$, we say $X_n$ is at most of order in probability $g_n$ and write

$$X_n = O_p(g_n).$$

Let $X_n$ is a $k$ dimensional random variable. If for every $\varepsilon > 0$, there exists a positive real number $M_\varepsilon$ such that

$$P\{|X_{jn}| \geq M_\varepsilon g_n\} \leq \varepsilon, \quad j = 1, 2, \ldots, k,$$

for all $n$, then we say $X_n$ is at most of order in probability $g_n$ and write

$$X_n = O_p(g_n).$$
Let $B_n$ be a $k \times r$ matrix of random variables. If for every $\varepsilon > 0$, there exists a positive real number $M_\varepsilon$ such that
\[ P\{|b_{ijn}| \geq M_\varepsilon g_n\} \leq \varepsilon, \quad i = 1, 2, \ldots, k, \; j = 1, 2, \ldots, r, \]
for all $n$, where $b_{ijn}$ are the elements of $B_n$. Then we say $B_n$ is at most of order in probability $g_n$ and write
\[ B_n = O_p(g_n). \]

We use $o_p$ to denote smaller order in probability. If
\[ p \lim g_n^{-1} X_n = 0, \]
we say $X_n$ is of smaller order in probability than $g_n$ and write
\[ X_n = o_p(g_n). \]

Let $X_n$ is a $k$ dimensional random variable. If for every $\varepsilon > 0$ and $\delta > 0$, there exists an $N$ such that for all $n > N$,
\[ P\{|X_{jn}| \geq \varepsilon g_n\} < \delta, \quad j = 1, 2, \ldots, k, \]
then we say $X_n$ is of smaller order in probability than $g_n$ and write
\[ X_n = o_p(g_n). \]

Let $B_n$ be a $k \times r$ matrix of random variables. If for every $\varepsilon > 0$ and $\delta > 0$, there exists an $N$ such that for all $n > N$,
\[ P\{|b_{ijn}| \geq \varepsilon g_n\} < \delta, \quad i = 1, 2, \ldots, k, \; j = 1, 2, \ldots, r, \]
then we say $B_n$ is of smaller order in probability than $g_n$ and write
\[ B_n = o_p(g_n). \]
Given a sequence of finite populations $\mathcal{F}_N$, the estimator $\hat{\theta}$ is said to be design consistent for the finite population parameter $\theta_N$, if for any fixed $\varepsilon > 0$,
\[
\lim_{N \to \infty} \Pr\{|\hat{\theta} - \theta_N| > \varepsilon|\mathcal{F}_N\} = 0.
\]
This notation indicates that given the fixed sequence of finite populations, the probability depends only on the sample design. (Fuller, 2006, pp.42)

1.4 Review of Estimation Approaches

1.4.1 Ordinary Least Squares Estimator

On the basis of model (1.3), the ordinary least squares (OLS) estimator of $\beta$ is
\[
\hat{\beta}_{ols} = \left(\sum_{i \in A} x_i x_i'\right)^{-1} \sum_{i \in A} x_i y_i = (X'X)^{-1} X'y, \tag{1.4}
\]
with conditional covariance matrix
\[
V\{\hat{\beta}_{ols}|X\} = (X'X)^{-1} \sigma^2.
\]
An estimator of the conditional variance of $\hat{\beta}_{ols}$ is
\[
\hat{V}\{\hat{\beta}_{ols}\} = (X'X)^{-1} \hat{\sigma}^2_{ols}; \tag{1.5}
\]
where
\[
\hat{\sigma}^2_{ols} = (n-k)^{-1} \sum_{i \in A} \hat{e}_{i,ols}^2,
\]
k is the dimension of $x_i$ and $\hat{e}_{i,ols} = y_i - x_i'\hat{\beta}_{ols}$. For fixed $X$, the OLS estimator is the best linear unbiased estimator of the superpopulation parameter $\beta$, given model (1.3).

Assume now that a sample is drawn from the population using a sample design $p(A)$ with associated $\pi_i$'s, where $\pi_i$ is the inclusion probability, or the selection probability
for element $i$. The inclusion probability is the probability that element $i$ is selected into the sample.

The conditional expected value of $X'y$ is

$$E\{X'y|F\} = E\left\{ \sum_{i \in A} x_i y_i | F \right\} = \sum_{i \in U} x_i \pi_i y_i = \sum_{i \in U} x_i \pi_i x'_i + \sum_{i \in U} x_i \pi_i e_i,$$

and the conditional expected value of $X'X$ is

$$E\{X'X|F\} = E\left\{ \sum_{i \in A} x_ix'_i | F \right\} = \sum_{i \in U} x_i \pi_i x'_i.$$

Under the moment assumption in Theorem 1 in Section 2.1.2, it can be proven that

$$\hat{\beta}_{ols} - \beta_{D_\pi,N|F} = \left( n^{-1} \sum_{i \in U} x_i \pi_i x'_i \right)^{-1} \left[ n^{-1} \sum_{i \in A} x_i (y_i - x'_i \beta_{D_\pi,N}) \right] + O_p(n^{-1}), \quad (1.6)$$

where

$$\beta_{D_\pi,N} = (X'_N D_{\pi,N} X_N)^{-1} X'_N D_{\pi,N} y_N,$$

and $D_{\pi,N} = diag(\pi_1, \pi_2, \ldots, \pi_N)$. Also

$$p \lim_{N \to \infty} \beta_{D_\pi,N} = \beta.$$

The probability limit of the OLS estimator is the weighted regression coefficient for the superpopulation where the weights are the selection probabilities.

The approximate bias of the OLS estimator is zero if $\pi_i x_i$ and $e_i$ are independent. If $\pi_i x_i$ and $e_i$ are correlated, then $E\{X'e|F\} \neq 0$ and the OLS estimator (1.4) is biased.

1.4.2 Probability Weighted Estimator

Consider a model for a sample of size $n$ selected from the finite population (1.2) as

$$y = X\beta + e,$$  \quad (1.7)
where \( y \) is the \( n \) dimensional column vector of observations on \( y \), \( X \) is the \( n \times k \) matrix of observations on the explanatory variables, and \( e \) is the \( n \) dimensional error vector. Assume that the sample is selected with unequal probabilities \( \pi_i \).

Under unequal probability sampling, a common procedure to account for possible sampling effects is the probability weighted (PW) estimator. The PW estimator, constructed with the inverses of the selection probabilities, is

\[
\hat{\beta}_{PW} = \left( \sum_{i \in A} x_i \pi_i^{-1} x_i' \right)^{-1} \sum_{i \in A} x_i \pi_i^{-1} y_i
\]

\[
= (X'WX)^{-1} X'Wy.
\]

(1.8)

where \( W = \text{diag}(\pi_1^{-1}, \pi_2^{-1}, \ldots, \pi_n^{-1}) =: \text{diag}(w_1, w_2, \ldots, w_n) \). We call \( w_i \) the sampling weight, which is the inverse of the selection probability \( \pi_i \). The sampling weight \( w_i \) can be viewed as the number of units in the population represented by the sample observation \( y_i \).

Under the moment assumption in Theorem 1 in Section 2.1.2, it can be proven that

\[
\hat{\beta}_{PW} - \beta_N | F = N^{-1} \sum_{i \in U} x_i x_i' \left[ N^{-1} \sum_{i \in A} x_i \pi_i^{-1} (y_i - x_i' \beta_N) \right] + O_p(n^{-1}),
\]

(1.9)

where

\[
\beta_N = (X_N'X_N)^{-1} X_N' y_N.
\]

The coefficient \( \beta_N \) is the ordinary least squares regression of \( y_N \) on \( X_N \) in the population. It can be proven that

\[
\hat{\beta}_{PW} - \beta_N | F = O_p(n^{-1/2}),
\]

(1.10)

and \( \hat{\beta}_{PW} \) is design consistent for the finite population parameter \( \beta_N \).

Under the model,

\[
E \left\{ n^{-1} \sum_{i \in A} x_i \pi_i^{-1} e_i \right\} = E \left\{ E \left\{ n^{-1} \sum_{i \in A} x_i \pi_i^{-1} e_i | F \right\} \right\}
\]

\[
= E \left\{ n^{-1} \sum_{i \in U} x_i e_i \right\}
\]

\[
= 0,
\]
and $V\{\beta_N - \beta\} = O(N^{-1})$. Then the probability weighted regression coefficient $\hat{\beta}_{PW}$
is a consistent estimator of the superpopulation parameter $\beta$.

If $e$ is independent of $\pi$, where $\pi = (\pi_1, \pi_2, \ldots, \pi_n)$, $e = (e_1, e_2, \ldots, e_n)$ and $e \sim (0, I\sigma^2)$, then the conditional covariance matrix of $\hat{\beta}_{PW} - \beta$ under model (1.7) is

$$V\{(\hat{\beta}_{PW} - \beta) | X\} = (X'WX)^{-1} X'WWX (X'WX)^{-1}\sigma^2. \quad (1.11)$$

If the selection is such that $y_i\pi_i^{-1}$ is uncorrelated with $y_j\pi_j^{-1}$ for $i \neq j$, an estimated covariance matrix of $\hat{\beta}_{PW}$ is

$$\hat{V}\{\hat{\beta}_{PW}\} = (X'WX)^{-1} X'\hat{D}_{ee,PW}WX (X'WX)^{-1}, \quad (1.12)$$

where

$$\hat{D}_{ee,PW} = diag(\hat{e}_{1,PW}^2, \hat{e}_{2,PW}^2, \ldots, \hat{e}_{n,PW}^2)$$

and $\hat{e}_{i,PW} = y_i - x_i'\hat{\beta}_{PW}$. In most cases the variance of the PW estimator is larger than the variance of the OLS estimator.

### 1.4.3 Pfeffermann-Sverchkov Estimator

When the sample selection probabilities are correlated with the model response variables after conditioning on the auxiliary variables, the sampling mechanism is called informative. Sugden and Smith (1984) examine the role of the sample selection mechanism in a model-based approach to finite population inference. Krieger and Pfeffermann (1992) and Pfeffermann (1993) discuss the notions of the informative sampling design based on the distribution of population measurements and the distribution of sample measurements.

Pfeffermann (1993) provides an example of an informative design. Suppose $(y_i, x_i)$ are independent draws from a bivariate normal distribution $N_2(\mu, \Sigma)$. Suppose $\{(y_i, x_i), i = 1, \ldots, n\}$ are observed for a sample of size $n$, and we want to estimate the population mean of $y_i$, $\mu_y = E\{y\}$. If the sample is selected by simple random sampling with
replacement, then the simple sample mean $\bar{y}$ is unbiased for $\mu_y$. The sample selection scheme can be ignored in this case in the inference process. However, if the sample is selected with probabilities proportional to $x_i$ with replacement, such that at each draw $k = 1, \ldots, n$, $P(i \in s) \propto x_i / \sum_{i=1}^{N} x_i$. If $\text{Corr}(Y, X) > 0$ and $P(y_i > \mu_y | i \in s) > 0.5$, then the distribution of $y_i$'s in the sample is different from that in the population and $E\{\bar{y}\} > \mu_y$. Ignoring the sampling scheme and estimating $\mu_y$ by $\bar{y}$ is misleading in this case.

When the selection probabilities are related to the values of the response variable, the empirical sample distribution is not consistent with the distribution of the population measurements, and the selection effects need to be accounted for in the inference process. The OLS estimator that ignores the sample selection process can yield large biases.

Skinner (1994) proposes an approach of extracting the population model from models fitted to the sample data. He showed two important propositions about relationships between the population distribution and the sample distribution. The first proposition shows that given the sampling weights, the conditional sample distributions of $y_i$ and $x_i$ are identical to the conditional population distributions of $y_i$ and $x_i$. The second proposition shows that the population of the sampling weights can be obtained by weighting the sample distribution of the sampling weights. Skinner presented a Monte Carlo study to compare the proposed procedure with the OLS estimator and the PW estimator based on the empirical bias, variance and mean squared error. Pfeffermann et al (1998) propose a general method of inference for the population distribution under the informative sampling that consists of approximating the parametric sample distribution. They showed how the sampling distribution may be derived from the population distribution using the first order selection probabilities for sample units. Pfeffermann and Sverchkov (1999) propose two new classes of estimators for regression models fitted to survey data. The proposed estimators account for the effect of informative sampling schemes and are derived from relationships between the popu-
lation distribution and the sample distribution. The first class consists of estimators obtained by extracting the sample distribution as a function of the population distribution and the sample selection probabilities and applying maximum likelihood theory to the sample distribution. The second class consists of estimators obtained by using relationships between the moments of the two distributions. The basis for the second class estimators is that the population regression \( E\{y_i|x_i\} \) can be obtained by using the sample regression \( y_i w_i \) and \( w_i \) on \( x_i \). The proposed estimator of the regression coefficient \( \beta \) from the second class is called the Pfeffermann-Sverchkov (PS) estimator in Chapter 2.

**1.4.4 Instrumental Variable Estimator**

Under the superpopulation model

\[
y_i = x_i' \beta + e_i,
\]

(1.13)

assume some members of \( x_i \) are not independent of \( e_i \). The crucial assumption for consistency is that \( x_i \) is independent of \( e_i \) in the superpopulation. If a variable is independent of the error term, this variable is called an exogenous variable. If an explanatory variable is correlated with the error term, this explanatory variable is sometimes called an endogenous explanatory variable. It is known that the presence of errors of measurement in the explanatory variable and the presence of endogenous explanatory variables in the regression model make the OLS estimator inconsistent and biased. For such cases, additional information is needed to obtain consistent parameter estimators.

Assume some additional variables, denoted by \( r_i \), are available with the superpopulation properties

\[
E\{r_i e_i\} = 0
\]

(1.14)

\[
|E\{x_i' r_i r_i' x_i\}| \neq 0,
\]

(1.15)
where $|C|$ is the determinant of the matrix $C$. Variables satisfying (1.14) and (1.15) are called the superpopulation instrumental variables or instruments. Thus, an instrumental variable (IV) must have two properties: (1) it must be uncorrelated with the error term of the structural equation; (2) it must be correlated with the endogenous explanatory variable. One method of instrumental variable estimation is called two-stage least squares (2SLS). For details see Wooldridge (2000).

The method of instrumental variables has been used for more than sixty years. In the 1940s the IV method was introduced for use in the Errors in Variables Model. See Reiersøl (1941, 1945). Geary (1947) shows that in certain cases consistent estimators may be obtained by the use of instrumental variables. Durbin (1954) reviews the IV approach to the problem of finding a consistent estimator of the regression coefficient. Sargan (1958) applies the IV method to a more general case and discusses the effect of increasing the number of instrumental variables. Sargan’s (1958) work and the instrumental variable character of two-stage least squares (2SLS) have made IV estimation widely used. The 2SLS method yields consistent estimates when one or more explanatory variables are endogenous in a regression model. The trade-off between bias and variance in the choice between the OLS estimator and the 2SLS estimator was considered in a Monte Carlo study by Summers (1965). Richardson and Wu (1971) analytically compare properties of the distribution function of the OLS and 2SLS estimators of structural coefficients in a simultaneous equation model that includes two endogenous variables. Richardson and Wu showed that the distribution function of the OLS estimator of the coefficient of the endogenous variable has the same form as the 2SLS estimator distribution, and compared the biases and mean squared errors of the estimators. Feldstein (1974) suggests and evaluates alternative procedures for balancing the loss of efficiency in the IV estimation against the potential gain of reduced bias. He considered two types of estimators: (1) a linear combination of the OLS and IV estimators and (2) a method of choosing between the OLS and IV estimators on the basis of sample information. Carter
and Fuller (1980) compare a modified IV estimator with randomly weighted average estimators of the type considered by Huntsberger (1955) and Feldstein (1974). Carter and Fuller study showed that the randomly weighted average estimators constructed for the IV estimator and the OLS estimator display the same type of behavior as the randomly weighted average estimators of two means studied by Mosteller (1948) and Huntsberger (1955). Aldrich (1993) presents a detailed study of the work of Reiersøl and Geary in the 1940s to explain the “idea of instrumental variables.”

1.4.5 A Test for Importance of Weights

In practice, it is often that not all the design variables are known for the whole population or that there are too many variables for all to be incorporated in the analysis. Not including all the design variables does not necessarily imply that the inference is biased and Sugden and Smith (1984) indicate that incorporating partial design information in the model can be sufficient for analytic inference about model parameters. A natural question arising from this topic is how to test that the design can be ignored in estimation, given the available design information.

The studies reviewed on this important aspect of the modeling process are mostly in the area of regression analysis. The ignorability of the design is tested by testing the significance of the difference between the OLS estimator $\hat{\beta}_{ols}$ and the weighted estimator $\hat{\beta}_{PW}$ under a working model which assumes that the design is ignorable.

DuMouchel and Duncan (1983) show that the difference between the weighted and unweighted estimates can be used as an aid in choosing the appropriate model and hence the appropriate estimator. The test is based on the difference $\hat{\Delta} = \hat{\beta}_{PW} - \hat{\beta}_{ols}$ with the null hypothesis that $H_0 : \Delta \equiv E\{\Delta\} = 0$. The test statistic is

$$\lambda = \hat{\Delta}' \hat{V}(\hat{\Delta})^{-1} \hat{\Delta},$$

(1.16)

where $\hat{V}(\hat{\Delta})$ is an estimator of the covariance matrix of $\hat{\Delta}$. 
The use of $\lambda$ of (1.16) for testing the importance of weights illustrates an important role for the sampling weights in the modeling process. The PW estimator is design consistent estimator for the population parameter $\beta_N$. If the sampling design is ignorable, the OLS estimator is likewise consistent for $\beta_N$. However, when the ignorability conditions are not satisfied, the OLS estimator is no longer consistent for $\beta_N$ and two estimators converge to different limits.

Fuller (1984) considers the case of a cluster sample within strata and estimated the covariance matrix $V(\hat{\Delta})$ by estimating the corresponding randomization covariance matrix. The resulting test statistic has an approximate $F$ distribution under $H_0$ with $k$ and $(n - 2k - L)$ degrees of freedom, where $L$ is the number of strata. The use of the randomization distribution to estimate $V(\hat{\Delta})$ is more robust than DuMouchel and Duncan’s approach because it does not depend on the regression model assumption.

Pfeffermann and Sverchkov (1999) suggest a formal test for testing sampling ignorability by using the moments relationships between the population distribution of regression residuals and the sample distribution of regression residuals. Let $\varepsilon_i = y_i - E\{y_i|\mathbf{x}_i\}$ denote the error term associated with unit $i$. Classical test procedures for comparing two distributions are not applicable since no observations are available for the population distribution of the residuals. However, under general conditions, the set of all moments of a distribution determine the distribution, provided the moments exist. Thus the null hypothesis is $H_0: E\{\varepsilon_i^k\} = E_s\{\varepsilon_i^k\}, k = 1, 2, \ldots$, where $E_s$ is the expectation under the sample distribution. The cumulative distribution function of sample $y_i$’s is defined as

$$P\{y_i \leq b|y_i \in A\} = F_s(b), \quad (1.17)$$

and

$$E_s\{y_i\} = \int y_i dF_s. \quad (1.18)$$

Pfeffermann and Sverchkov showed the moments relationship between the sample and
population pdf’s,

\[ E\{ \mathbf{u}_i | \mathbf{v}_i \} = [E_s\{ w_i | \mathbf{v}_i \}]^{-1} E_s\{ \mathbf{u}_i w_i | \mathbf{v}_i \} \]  \hspace{1cm} (1.19)

for any pair of vector random variables \(( \mathbf{u}_i, \mathbf{v}_i )\), where \( w_i = \pi_i^{-1} \) is the sampling weight for unit \( i \). By the moments relationship \( E\{ \mathbf{v}_i^{k} \} = [E_s\{ w_i \}]^{-1} E_s\{ \mathbf{v}_i^{k} w_i \} \) which is a special case of (1.19), an equivalent set of hypotheses is

\[ H_{0k} : Corr_s(\mathbf{v}_i^{k}, \mathbf{w}_i) = 0, \quad k = 1, 2, \ldots \]

where \( Corr_s \) is the correlation under the sample distribution. In Pfeffermann and Sverchkov (1999)’s simulation study, Pfeffermann and Sverchkov use as test statistic a standardized form of the Fisher transformation of the correlation coefficient \( FT(k) = (1/2) \log[(1 + r_k)/(1 - r_k)] \), and

\[ FTS(k) = FT(k)/\hat{SD}(FT(k)), \]

where \( r_k \) is the empirical correlation \( \hat{Corr}(\mathbf{v}_i^{k}, \mathbf{w}_i) \), \( \mathbf{v}_i^{k} = (y_i - \mathbf{x}_i^{k} \beta)^k \) and \( \hat{SD}(FT(k)) \) is the bootstrap standard deviation of \( FT(k) \). The test statistic has an asymptotic normal distribution with mean zero.

### 1.4.6 Preliminary Testing Procedures

The motivation for the preliminary testing (pretest) procedure is to accept bias in return for reduced variance. Pretest estimators are a class of estimators that make a trade for smaller variance at the risk of bias. The pretest procedure is characterized by a test statistic, \( T \), calculated from the data set. The test \( T \) determines the estimation method. If \( T \) is statistically significant, a given procedure will be used to estimate a parameter. Otherwise an alternative procedure will be used for calculating the estimator.

The general idea of using a pretest to determine an estimation procedure is discussed by Bancroft (1944), Mosteller (1948) and Huntsberger (1955). Bancroft (1944) proposes...
pretest procedures for testing homogeneity of variances and testing of a regression coefficient. Bancroft’s paper does not provide a clear-cut prescription of when or whether to pretest. Mosteller (1948) discusses a simple problem concerning the pooling of data and presents several ways of pooling data from two samples to estimate the mean of the population of one of them. Mosteller pointed out that if the difference between the true means can be thought of as normally distributed from sample to sample, pooling with unequal weights is preferable. A generalization of the sometimes-pool procedure for pooling two estimators which is based on a pretest was described by Huntsberger (1955). He compared the efficiencies of the generalized weighting procedure and of the sometimes-pool procedure for the special case where the estimators are normally distributed.

To formulate the sometimes-pool idea, we follow Huntsberger (1955). Suppose we have a random sample with two unknown parameters $\theta_1$ and $\theta_2$. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be estimators for $\theta_1$ and $\theta_2$. In general, when $\theta_1 = \theta_2$, a pooled estimator $g(\hat{\theta}_1, \hat{\theta}_2)$ will provide a better estimator for $\theta_1$ than $\hat{\theta}_1$ alone. When it’s unclear whether or not $\theta_1$ is equal to $\theta_2$, the pooled estimator may still provide some gains, but may lose when the two parameters are very different. Let our pretest statistic $T$, be the statistic for testing the null hypothesis $H_0 : \theta_1 = \theta_2$ against the alternative hypothesis $H_\alpha : \theta_1 \neq \theta_2$. An estimator can be formed for $\theta_1$ using the function

$$W(T) = \phi(T)\hat{\theta}_1 + [1 - \phi(T)]g(\hat{\theta}_1, \hat{\theta}_2),$$  

where $\phi(T)$ is an indicator defined as

$$\phi(T) = \begin{cases} 
0 & \text{if } T \subset A_\alpha \\
1 & \text{if } T \subset R_\alpha, 
\end{cases}$$ 

where $A_\alpha$ and $R_\alpha$ are the acceptance and rejection regions for the test of the null hypothesis that $\theta_1 = \theta_2$ with the significance level $\alpha$. 
Rao (1966) arrives at a similar conclusion to Huntsberger (1955). Under a probability proportional to size (pps) sampling design, he proposed two estimators, a Horvitz-Thompson (H-T) estimator, and an alternative estimator. The H-T estimator is a weighted estimator. The alternative estimator is an unweighted estimator and of the form of the simple sample mean. Rao stated that there is a criterion to choose between the two estimators, the correlation between a characteristic of interest \( y \) and the selection probabilities. If the characteristics are poorly correlated with the selection probabilities in the pps sampling design, the alternative estimator may be used. For other characteristics, the weighted estimator should be used. The alternative estimator has smaller mean squared error than the H-T estimator and the bias is relatively small compared to the standard error when characteristics are poorly correlated with the selection probabilities in pps sampling design. Rao didn’t propose a test procedure for testing if a characteristic of interest \( y \) and the selection probabilities are correlated.

Bock, Yancey and Judge (1973) develop the properties of the pretest estimator for the general model and determine the characteristics of the risk function of the pretest estimator under the squared error loss criterion. The choice of the level of significance for the test was discussed in the paper. Cohen (1974) gives necessary and sufficient conditions for the procedures based on the pretest of significant to be admissible. As stated by Cohen (1974), the pretest procedure is sort of a compromise between a Bayesian procedure and the usual procedure. Inference based on the pretest procedure requires, in general, less prior knowledge on the part of research workers than the use of Bayesian inference procedures. With limited knowledge, the researchers tend to check the assumptions by using pretest procedures. Wallace (1977) reviews the properties of various pretest estimators both in the general case (multiple restrictions) and in the Bancroft case (a single restriction) and works on optimal pretest critical values. Bancroft and Han (1977) comply a bibliography containing references on the topic about the pretest procedure. Grossman (1986) suggests a compromise sampling strategy between a large supposedly
unpoolable sample and a smaller supposedly poolable sample based on pretest procedures in social experiments. Gregoire, Arabatzis and Reynolds (1992) provide a pretest procedure for the intercept in a simple linear regression model. Magnus and Durbin (1999) use a pretest procedure for estimation of regression coefficients of interest when other regression coefficients are a vector of nuisance parameters.

In the context of the pretest procedure based on the IV estimator, Sargan (1958) suggests obtaining the confidence interval for the IV estimator and notes that if the OLS estimator lies outside this interval, it is “probably significantly biased.” Although there was no formal test proposed in Sargan’s paper, Sargan’s remarks suggest that if the absolute difference between the OLS estimator and the IV estimator is greater than the standard error of the IV estimator, infer that the OLS estimator is biased and use the IV estimator; if not, use the OLS estimator. This procedure is a pretest estimator.
CHAPTER 2 Design Consistent Estimation

2.1 H Estimator

2.1.1 Motivation

Assume that the ordinary least squares (OLS) estimator is biased. In such cases it is necessary to incorporate the sampling weights into the analysis. One approach is to use the probability weighted (PW) estimator. The PW estimator is consistent for $\beta$, but it can be inefficient due to unequal probabilities not being proportional to the conditional variance of $y$.

An alternative approach to constructing design consistent estimators that are more efficient under the model (1.7) than the PW estimators is to find procedures to scale the sampling weights to near one or to the conditional variance of $y$.

Consider an estimator of the form

$$\hat{\beta}_H = \left(\sum_{i \in A} x_i \pi_i^{-1} h_i x_i'\right)^{-1} \sum_{i \in A} x_i \pi_i^{-1} h_i y_i$$

where $H$ is a diagonal matrix with diagonal elements $h_i$, $h_i$ is defined for all $i \in U$ and $h_i$ is independent of $e_i$. We call the estimator $\hat{\beta}_H$ the H estimator.

2.1.2 Central Limit Theorem

In this section, we show that the H estimator is consistent for the population parameter under mild assumptions and has a limiting normal distribution. We begin with

**Lemma 1.** Let \( \{V_n\} \) be a sequence of random variables in \( \mathbb{R}^k \) such that, for some function \( h \), as \( n \to \infty \),

\[
h(V_1, \ldots, V_n) \xrightarrow{\mathcal{L}} \Gamma,
\]

where \( \Gamma \) has a distribution function \( G \). If \( \{L_n\} \) is a sequence of random variables in \( \mathbb{R}^k \) such that

\[
P\{L_n - h(V_1, \ldots, V_n) \leq s|V_1, \ldots, V_n\} \to F(s)
\]

almost surely for all \( s \in \mathbb{R}^k \), where \( F \) is a continuous distribution function, \( | \) represents conditional upon, and \( \leq \) is taken to mean jointly less than elementwise. Then

\[
P(L_n \leq t) \to (G * F)(t),
\]

for all \( t \in \mathbb{R}^k \), where “*” denotes convolution.

Lemma 1 is in terms of generic CDFs \( G \) and \( F \). Lemma 2 is a special case of Lemma 1. Lemma 2 is an application of Lemma 1 to normal CDFs. The proof of Lemma 2 is given in Legg (2006).

**Lemma 2.** Let \( \{F_N\} \) be a sequence of finite populations and let \( \theta_N \) be a function on \( \mathbb{R}^k \) of the elements of \( F_N \) such that

\[
N^{1/2}(\theta_N - \theta) \xrightarrow{\mathcal{L}} N_k(0, V_{11}).
\]

Let a design and an estimator, \( \hat{\theta}_N \), and a sequence of conditional variance matrices \( V_{22,N} \) be such that

\[
N^{1/2}(\hat{\theta}_N - \theta_N)|F_N \xrightarrow{\mathcal{L}} N_k(0, V_{22}) \quad \text{a.s.},
\]
\[
\lim_{N \to \infty} V_{22,N} = V_{22} \quad \text{a.s.,} \quad (2.7)
\]

where \( V_{11} + V_{22,N} \) is positive definite for all \( N \). Then

\[
N^{1/2}(V_{11} + V_{22,N})^{-1/2}(\hat{\theta}_N - \theta) \xrightarrow{\mathcal{L}} N_k(0, I_k), \quad (2.8)
\]

where \( I_k \) is the \( k \times k \) identity matrix.

We prove the main result of this section in the following theorem.

**Theorem 1.** Let \( \{(y_i, x_i, h_i)\} \) be a sequence of independent identically distributed random variables with 9-th moment. Let \( \{U_N, \mathcal{F}_N : N = k + 3, k + 4, \ldots\} \) be a sequence of finite populations, where \( U_N \) is the set of indices identifying the elements and \( \mathcal{F}_N = ((y_1, x_1, h_1), \ldots, (y_N, x_N, h_N)) \). In the superpopulation \( y_i \) is related to \( x_i \) through a regression model,

\[
y_i = x_i' \beta + e_i,
\]

\( e_i \sim \text{ind}(0, \sigma^2) \).

Assume \( (x_i, h_i) \) is independent of \( e_i \). Assume \( h_i \) are positive for all \( i \). Let \( z_i = (y_i, x_i) \),

\[
M_{\mathcal{Z}H, N} = N^{-1} Z_N' H_N Z_N \quad (2.10)
\]

and

\[
M_{\mathcal{Z}H} = E\{M_{\mathcal{Z}H, N}\}, \quad (2.11)
\]

where \( H_N = \text{diag}(h_1, h_2, \ldots, h_N) \). Assume \( M_{zz} = E\{z_i' z_i\} \) and \( M_{\mathcal{Z}H} \) are positive definite.

Let \( Z = (z_1, z_2, \ldots, z_{n_N})' \), where we index the sample elements from one to \( n_N \). Let

\[
\hat{M}_{\mathcal{Z}H} = N^{-1} Z' W H Z, \quad (2.12)
\]
where $W = \text{diag}(\pi_1^{-1}, \pi_2^{-1}, \ldots, \pi_n^{-1}) =: \text{diag}(w_1, w_2, \ldots, w_N)$ and $H = \text{diag}(h_1, h_2, \ldots, h_N)$.

Assume the sequence of sample designs is such that for any $z$ with 3-rd moments
\[
\lim_{N \to \infty} n_N V\{\bar{z}_{HT} - \bar{z}_N|F_N\} = V_{zz,\infty} \quad \text{a.s.,} 
\]
and
\[
[V\{\bar{z}_{HT} - \bar{z}_N|F_N\}]^{-1/2}(\bar{z}_{HT} - \bar{z}_N)|F_N \xrightarrow{\mathcal{L}} N(0, I) \quad \text{a.s.}
\]
where
\[
\bar{z}_{HT} = N^{-1} \sum_{i \in A} \pi_i^{-1} z_i,
\]
$\bar{z}_N$ is the finite population mean of $z$, and $V\{\bar{z}_{HT} - \bar{z}_N|F_N\}$ and $V_{zz,\infty}$ are positive definite.

Assume $\lim_{N \to \infty} f_N = f$ a.s., where $f_N = N^{-1} n_N$ and $0 \leq f < 1$. Let $\hat{V}\{\bar{z}_{HT}\}$ be the Horvitz-Thompson variance estimator of $V\{\bar{z}_{HT}|F_N\}$, and assume
\[
\hat{V}\{\bar{z}_{HT}\} - V\{\bar{z}_{HT}|F_N\} = o_p(n_N^{-1})
\]
for any $z$ with 3-rd moments.

Let $\hat{\beta}_{HT}$ be defined by (2.1). Then
\[
\hat{\beta}_{HT} - \beta = M_{XXH}^{-1} \bar{b}_{HT} + O_p(n_N^{-1}),
\]
and
\[
\hat{\beta}_H - \beta_N = M_{XHX,N}^{-1} (\bar{b}_{HT} - \bar{b}_N) + O_p(n_N^{-1}),
\]
where
\[
\beta_N = (X_N' H_N X_N)^{-1} X_N' H_N y_N =: M_{XHX,N}^{-1} M_{XHY,N},
\]
\[
\bar{b}_N = N^{-1} \sum_{i \in U} b_i,
\]
\[
\bar{b}_{HT} = N^{-1} \sum_{i \in A} \pi_i^{-1} b_i,
\]
and \( b_i = x_i h_i e_i \).

Then
\[
n_{N}^{1/2}(V_{11,N} + f_{N}V_{22})^{-1/2}(\hat{\beta}_H - \beta) \xrightarrow{\mathcal{L}} N(0, I),
\]
(2.19)
as \( N \to \infty \), where \( V_{11,N} = n_{N}M^{-1}_{XHX,N}V(B_{HT}|F_{N})M^{-1}_{XHX,N} \), \( V_{22} = M^{-1}_{XHX} \Sigma_{bb}M^{-1}_{XHX} \) and \( \Sigma_{bb} = E\{b_i b'_i\} \).

Let \( \hat{\Sigma}_{bb} = \Sigma_{bb} + o_p(1) \). Then the estimated variance is
\[
\hat{V}\{\hat{\beta}_H\} = \hat{M}^{-1}_{XHX}(\hat{V}\{B_{HT}\} + N^{-1}\hat{\Sigma}_{bb})\hat{M}^{-1}_{XHX}
\]
(2.20)
where \( \hat{V}\{B_{HT}\} \) is the Horvitz-Thompson estimated sampling variance of \( \tilde{b}_{HT} \) calculated with \( \tilde{b}_i = x_i h_i \hat{e}_i \) and \( \hat{e}_i = y_i - x_i' \hat{\beta}_H \).

Proof. By the design,
\[
E\left\{N^{-1}\sum_{i\in A} x_i w_i h_i x'_i | F_N \right\} = N^{-1}\sum_{i\in U} x_i h_i x'_i,
\]
By moment assumptions and assumption (2.13),
\[
N^{-1}\sum_{i\in A} x_i w_i h_i x'_i - N^{-1}\sum_{i\in U} x_i h_i x'_i | F_N = \hat{M}^{-1}_{XHX} - M^{-1}_{XHX,N} | F_N = O_p(n_{N}^{-1/2}) \quad a.s.
\]
From a Taylor expansion,
\[
\hat{M}^{-1}_{XHX} - M^{-1}_{XHX,N} | F_N = O_p(n_{N}^{-1/2}) \quad a.s.
\]
Similary
\[
N^{-1}\sum_{i\in A} x_i w_i h_i e_i - N^{-1}\sum_{i\in U} x_i h_i e_i | F_N = \hat{M}^{-1}_{XHe} - M^{-1}_{XHe,N} | F_N = O_p(n_{N}^{-1/2}) \quad a.s.
\]
Under the model (2.9), by assumption, \( e_N \) is independent of \( (X_N, H_N) \). Thus
\[
M_{XHe} = E\{M_{XHe,N}\} = 0,
\]
and
\[ \hat{M}_{XHe} = O_p(n_N^{-1/2}). \]

By moment assumptions,
\[ M_{XHX,N} - M_{XHX} = O_p(N^{-1/2}). \]

Thus
\[
\hat{\beta}_H - \beta = (X'WHX)^{-1} X'WH(y - X\beta) = \hat{M}_{XHX,N}^{-1} \sum_{i \in A} x_i w_i h_i e_i = \hat{M}_{XHX,N}^{-1} M_{XHe} = M_{XHX,N}^{-1} \bar{b}_{HT} + O_p(n_N^{-1})
\]

and (2.16) is proven. Also
\[
\beta_N - \beta = (X'_N H_N X_N)^{-1} X'_N H_N (y_N - X_N \beta) = M_{XHX,N}^{-1} \sum_{i \in U} x_i h_i e_i = M_{XHX,N}^{-1} \bar{b}_N + O_p(N^{-1}).
\]

It follows that
\[
\hat{\beta}_H - \beta_N = (\hat{\beta}_H - \beta) - (\beta_N - \beta) = M_{XHX,N}^{-1} (\bar{b}_{HT} - \bar{b}_N) + O_p(n_N^{-1})
\]

and (2.17) is proven.

From variance assumption (2.13) and the normality assumption (2.14),
\[
n_N^{1/2}(\bar{b}_{HT} - \bar{b}_N)|\mathcal{F}_N \overset{L}{\to} N(0, V_{bb,\infty}) \quad a.s.,
\]
where $V_{bb, \infty}$ is defined by analogy to (2.13). Therefore

$$n_N^{1/2}(\hat{\beta}_H - \beta_N) |\mathcal{F}_N \xrightarrow{\mathcal{L}} N(0, V_{11}) \quad \text{a.s.},$$

(2.21)

where $V_{11} = M_{XH, X}^{-1} V_{bb, \infty} M_{XH, X}^{-1}$.

By the Central Limit Theorem for independent random variables and the moment assumptions

$$N^{1/2} \hat{\beta}_N \xrightarrow{\mathcal{L}} N(0, \Sigma_{bb}),$$

as $N \to \infty$. Then

$$N^{1/2}(\beta_N - \beta) \xrightarrow{\mathcal{L}} N(0, V_{22}),$$

(2.22)

as $N \to \infty$, where $V_{22}$ is defined in (2.19).

Applying Lemma 2, result (2.19) then follows from (2.21) and (2.22).

By the design,

$$\hat{V}\{\hat{b}_{HT}\} = O_p(n_N^{-1})$$

and

$$\hat{\Sigma}_{bb} = O_p(1),$$

then

$$\hat{V}\{\hat{b}_{HT}\} + N^{-1} \hat{\Sigma}_{bb} = O_p(n_N^{-1}).$$

(2.23)

By (2.23) and (2.15),

$$M_{XH, X}^{-1}(\hat{V}\{\hat{b}_{HT}\} + N^{-1} \hat{\Sigma}_{bb}) M_{XH, X}^{-1} = M_{XH, X, N}(\hat{V}\{\hat{b}_{HT}\} + N^{-1} \hat{\Sigma}_{bb}) M_{XH, X, N}^{-1} + O_p(n_N^{-3/2})$$

$$= M_{XH, X, N}(V\{\hat{b}_{HT}|\mathcal{F}_N\} + N^{-1} \Sigma_{bb} + o_p(n_N^{-1}))$$

$$+ O_p(n_N^{-3/2})$$

$$= M_{XH, X, N}(V\{\hat{b}_{HT}|\mathcal{F}_N\} + N^{-1} \Sigma_{bb}) M_{XH, X, N}^{-1}$$

$$+ o_p(n_N^{-1})$$
\[ \begin{align*}
&= M^{-1}_{XH,N} V \{ \tilde{b}_{HT} | \mathcal{F}_N \} M^{-1}_{XH,N} \\
&+ N^{-1} M^{-1}_{XH} \Sigma_{bb} M^{-1}_{XH} + o_p(n^{-1}_N) \\
&= n^{-1}_N V_{11,N} + N^{-1} V_{22} + o_p(n^{-1}_N) \\
&= n^{-1}_N (V_{11,N} + f_N V_{22}) + o_p(n^{-1}_N). \quad (2.24)
\]

By Theorem 2.2.1 (Fuller, 2006), we can replace \( e_i \) with \( \hat{e}_i \) in (2.24). Result (2.20) then follows.

Theorem 1 is a consequence of almost sure convergence assumptions. Under the regularity conditions of Theorem 1, \( \hat{\beta}_H \) is consistent for \( \beta \) for any \( H \) matrix that meets the moment assumptions in Theorem 1 and is independent of \( e \). The \( H \) matrix plays little role due to independence assumptions. The proof of Theorem 1 is a proof of consistency for the probability weighted regression estimator obtained by setting \( H = I \).

In order to illustrate how to construct an \( H \) estimator, an example of choosing \( H \) matrix is given. Assume Poisson sampling from a finite population generated as \( iid \) random variables and consider a model

\[ y_i = \beta x_i + e_i, \quad (2.25) \]

where \( e_i = [g(x_i)]^{1/2} a_i \), \( a_i \sim ind(0, 1) \) and is independent of \( x_i \), and \( g(\cdot) \) is a positive bounded function. In the superpopulation, \( x_i \) is independent of \( e_i \). Assume \( (x_i, \sigma^2_{e_i}, a_i) \) are \( iid \) random vectors, where \( \sigma^2_{e_i} = g(x_i) \) is the variance of \( e_i \). Under Poisson sampling, \( \pi_i^{-1} e_i \) is independent of \( \pi_j^{-1} e_j \), for all \( i \) and \( j \), and

\[ V \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i | \mathcal{F} \right\} = \sum_{i \in U} \pi_i^{-1} (1 - \pi_i) x_i^2 h_i^2 e_i^2. \quad (2.26) \]

Assume

\[ w_i = \pi_i^{-1} = k(x_i) + v_i, \]
where \( k(\cdot) \) is a positive bounded function and \( v_i \) is independent of \((x_i, e_i)\). We have

\[
E \left\{ \sum_{i \in A} \pi_i^{-1} x_i h_i e_i | \mathcal{F} \right\} = \sum_{i \in U} x_i h_i e_i. \tag{2.27}
\]

Under the independence assumptions, the unconditional variance \( V \{ \sum_{i \in A} \pi_i^{-1} x_i h_i e_i \} \) is

\[
V \left\{ \sum_{i \in A} \pi_i^{-1} x_i h_i e_i \right\} = E \left\{ V \left\{ \sum_{i \in A} \pi_i^{-1} x_i h_i e_i | \mathcal{F} \right\} \right\} + V \left\{ E \left\{ \sum_{i \in A} \pi_i^{-1} x_i h_i e_i | \mathcal{F} \right\} \right\}
\]

\[
= E \left\{ \sum_{i \in U} \pi_i^{-1} (1 - \pi_i) x_i^2 h_i^2 e_i^2 \right\} + V \left\{ \sum_{i \in U} x_i h_i e_i \right\}
\]

\[
= E \left\{ \sum_{i \in U} \pi_i^{-1} x_i^2 h_i^2 e_i^2 \right\} - E \left\{ \sum_{i \in U} x_i^2 h_i^2 e_i^2 \right\} + E \left\{ \sum_{i \in U} x_i^2 h_i^2 e_i^2 \right\}
\]

\[
= E \left\{ \sum_{i \in U} (k(x_i) + v_i) x_i^2 h_i^2 e_i^2 \right\}
\]

\[
= E \left\{ \sum_{i \in U} k(x_i) x_i^2 h_i^2 e_i^2 \right\} + E \left\{ \sum_{i \in U} v_i x_i^2 h_i^2 e_i^2 \right\}
\]

\[
= \sum_{i \in U} E \{ k(x_1) x_1^2 h_1^2 g(x_1) \}. \tag{2.28}
\]

Then

\[
V \{ \hat{\beta}_H - \beta \} \simeq (NE \{ x_1^2 h_1 \})^{-2} NE \{ k(x_1) x_1^2 h_1^2 g(x_1) \}. \tag{2.28}
\]

Therefore, if \( k(x_i) \) and \( g(x_i) \) are known, we choose

\[
h_i = k(x_i)^{-1} g(x_i)^{-1}, \quad i = 1, 2, \ldots, n. \tag{2.29}
\]

to minimize the unconditional variance (2.28). Also see Fuller (2006).

We give a consistent variance estimator for \( \hat{\beta}_H - \beta \) for Poisson sampling. The conditional variance of \( X'WHe \) for Poisson sampling from a finite population is

\[
V \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i | \mathcal{F} \right\} = \sum_{i \in U} \pi_i^{-1} (1 - \pi_i) x_i h_i^2 e_i^2 x'_i.
\]

The estimated conditional variance of \( V \{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i | \mathcal{F} \} \) is

\[
\hat{V} \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i | \mathcal{F} \right\} = \sum_{i \in A} \pi_i^{-2} (1 - \pi_i) x_i h_i^2 e_i^2 x'_i.
\]
Also

\[
E \left\{ \hat{V} \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \mid F \right\} \right\} = E \left\{ \sum_{i \in U} \pi_i^{-1} (1 - \pi_i) x_i h_i^2 e_i^2 x_i' \right\} = E \left\{ \sum_{i \in U} \pi_i^{-1} x_i h_i^2 e_i^2 x_i' \right\} - E \left\{ \sum_{i \in U} x_i h_i^2 e_i^2 x_i' \right\}
\]

and

\[
V \left\{ \hat{V} \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \mid F \right\} \right\} = \hat{V} \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \right\} = E \left\{ \sum_{i \in U} \pi_i^{-1} x_i h_i^2 e_i^2 x_i' \right\}.
\]

Therefore, the unconditional variance of \( \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \) is

\[
V \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \right\} = E \left\{ \sum_{i \in U} \pi_i^{-1} x_i h_i^2 e_i^2 x_i' \right\}.
\]

The estimated unconditional variance of \( V \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \right\} \) is

\[
\hat{V} \left\{ \sum_{i \in A} x_i \pi_i^{-1} h_i e_i \right\} = \sum_{i \in A} \pi_i^{-2} x_i h_i^2 x_i' \hat{e}_i^2,
\]

and an estimated covariance matrix of \( \hat{\beta}_H \) for Poisson sampling is

\[
\hat{V} \{ \hat{\beta}_H \} = n(n - k)^{-1} (X' WH X)^{-1} X' WH \hat{D}_{ee,H} HW X (X' WH X)^{-1}, \quad (2.30)
\]

where \( k \) is the dimension of \( x_i \),

\[
\hat{D}_{ee,H} = diag(\hat{e}_1^2, \hat{e}_2^2, \ldots, \hat{e}_n^2),
\]

\( \hat{e}_i = y_i - x'_i \hat{\beta}_H \). The estimated variance (2.30) is a consistent estimator of the variance of \( \hat{\beta}_H - \beta \).

The variance expression (2.30) can be used for a stratified sample selected from a population which is a simple random sample from a stratified superpopulation. For stratified sampling, the conditional variance of the stratified mean is

\[
V \{ \bar{y}_{st} - \bar{y}_N \mid F \} = \sum_{h=1}^{H} \sum_{i \in U_h} N^{-2} \pi_{hi}^{-1} (1 - \pi_{hi}) N_h (N_h - 1)^{-1} \hat{e}_{hi}^2,
\]
where $\pi_{hi} = n_h N_h^{-1}$ is the selection probability in the stratum $h$, $e_{hi} = y_i - \bar{y}_{Nh}$, $\bar{y}_N$ is the finite population mean and $\bar{y}_{Nh} = N_h^{-1} \sum_{i \in U_h} y_i$. The estimated conditional variance is

$$\hat{V}\{\bar{y}_{st} - \bar{y}_N | F\} = \sum_{h=1}^{H} \sum_{i \in A_h} N_h^{-2} \pi_{hi}^{-2} (1 - \pi_{hi}) n_h (n_h - 1)^{-1} \hat{e}_{hi}^2,$$

where $\hat{e}_{hi} = y_i - \bar{y}_{nh}$ and $\bar{y}_{nh} = n_h^{-1} \sum_{i \in A_h} y_i$. Thus the estimated unconditional variance is

$$\hat{V}\{\bar{y}_{st} - \bar{y}_N \} = \sum_{h=1}^{H} \sum_{i \in A_h} N_h^{-2} \pi_{hi}^{-2} n_h (n_h - 1)^{-1} \hat{e}_{hi}^2.$$  

For a general design, if a finite population correction can be ignored, a variance estimator of $\hat{\beta}_H$ is

$$\hat{V}\{\hat{\beta}_H | F\} = (X'WHX)^{-1} \hat{V}\{X'WHe\} (X'WHX)^{-1},$$

(2.31)

where $\hat{V}\{X'WHe\} = \hat{V}\{\sum_{i \in A} x_i \pi_{i}^{-1} h_i e_i\}$ is the Horvitz-Thompson estimator of the variance of the sum calculated with $x_i \pi_{i}^{-1} h_i e_i$ and $\hat{e}_i = y_i - x_i' \hat{\beta}_H$. Under assumptions given in Theorem 1, the variance estimator (2.31) is consistent for $V\{\hat{\beta}_H - \beta_N | F\}$.

### 2.2 Pfeffermann-Sverchkov Estimator

Pfeffermann and Sverchkov (1999) consider the regression model (1.7) and a design in which the $\pi_i$ may be a function of $x_i$ and $e_i$. Pfeffermann and Sverchkov proposed an estimator, that we call the Pfeffermann-Sverchkov (PS) estimator, obtained by utilizing information about the moments in the population. Estimation is a two-step procedure:

1. Calculate $\hat{w}_i$ by the regression of $w_i$ on known functions of $x_i$ using the sample measurements.
2. Compute

$$\hat{\beta}_{PS} = \arg \min_{\beta} \left\{ n^{-1} \sum_{i \in A} \hat{w}_i^{-1} w_i (y_i - x_i' \beta)^2 \right\}.$$

The $\hat{w}_i$ is an estimator of $E_s \{w_i | x_i\}$. The PS estimator $\hat{\beta}_{PS}$ of $\beta$ is calculated as

$$\hat{\beta}_{PS} = \left( \sum_{i \in A} q_i x_i x_i' \right)^{-1} \sum_{i \in A} q_i x_i y_i.$$
\[ (X'QX)^{-1} X'Qy, \] (2.32)

where

\[ Q = \text{diag}(q_1, q_2, \ldots, q_n), \]

\[ q_i = w_i \hat{w}_i^{-1}, \] and \( \hat{w}_i \) is the fitted value from the OLS regression of \( w_i \) on known functions of \( x_i \). The PS estimator is a version of the H estimator (2.1) with \( H = \hat{W}^{-1} = \text{diag}(\hat{w}_1^{-1}, \hat{w}_2^{-1}, \ldots, \hat{w}_i^{-1}) \).

If the model has constant error variances and the correlation between \( w_i \) and \( e_i^2 \) is modest, then \( E_s\{w_i|x_i\} \) will be highly correlated with \( V\{\pi_i^{-1}e_i\} \) and the PS estimator will perform well. If there is reasonable correlation between \( x_i \) and \( \pi_i^{-1}e_i^2 \), then the PS estimator can be used as the initial estimator to provide \( \hat{e}_i^2 \) to estimate \( h_i \) for constructing an H estimator.

The PW estimator and PS estimator coincide when \( \pi_i \)'s are independent of \( x_i \)'s such that \( E_s\{w_i|x_i\} = [E\{\pi_i|x_i\}]^{-1} = \text{constant} \). When \( w_i \) is a deterministic function of \( x_i \), \( q_i = 1 \) and \( \hat{\beta}_{PS} = \hat{\beta}_{ols} \). In general, \( \hat{w}_i \) is not a consistent estimator of the superpopulation expected value of \( w_i \) given \( x_i \). The variance from estimated \( E_s\{w_i|x_i\} \) may reduce the efficiency of the PS estimator. See Skinner (1994).

The PS estimator \( \hat{\beta}_{PS} \) is not strictly unbiased for \( \beta \), but under the assumptions of Theorem 1, it is consistent. We note that

\[
E\{N^{-1}X'Qe\} = E\left\{N^{-1} \sum_{i \in A} x_i q_i e_i \right\} \\
= E\left\{N^{-1} \sum_{i \in U} \pi_i (x_i w_i \hat{w}_i^{-1} e_i) \right\} \\
= E\left\{N^{-1} \sum_{i \in U} x_i \hat{w}_i^{-1} e_i \right\} \\
= 0. 
\]

The approximation follows from the fact that \( \hat{w}_i \) is an estimator of \( E\{w_i|x_i\} \) and de-
ependent on \( w_i \).

\[
E\{\hat{\beta}_{PS} - \beta\} = E\left\{ \left( N^{-1} \sum_{i \in U} x_i \hat{w}_i^{-1} x'_i \right)^{-1} N^{-1} \sum_{i \in U} x_i \hat{w}_i^{-1} e_i \right\}.
\]

for fixed \( \hat{w}_i^{-1} \). The lack of strict unbiasedness follows from the facts that the PS estimator is a ratio estimator and that \( e_i \) and \( q_i \) may be dependent when \( \pi_i \) depends on \( e_i \).

An estimated covariance matrix of \( \hat{\beta}_{PS} \) for Poisson sampling is

\[
\hat{V}\{\hat{\beta}_{PS}\} = n(n-k)^{-1} (X'QX)^{-1} X'Q \hat{D}_{ee,PS} QX (X'QX)^{-1},
\]

where \( k \) is the dimension of \( x_i \),

\[
\hat{D}_{ee,PS} = \text{diag}(\hat{e}^2_{1,PS}, \hat{e}^2_{2,PS}, \ldots, \hat{e}^2_{n,PS}),
\]

and \( \hat{e}_{i,PS} = y_i - x'_i \hat{\beta}_{i,PS} \).
CHAPTER 3 Instrumental Variable Estimation

3.1 Introduction

Consider a regression model

\[ y_i = \beta_0 + \beta_1 x_i + e_i, \]  
\[ e_i \sim (0, \sigma^2), \]

where \( \text{Cov}(x_i, e_i) \neq 0 \). In order to obtain consistent estimators of \( \beta_0 \) and \( \beta_1 \) under the model (3.1), we let \( z_i \) be an instrumental variable (IV) for \( x_i \).

The IV estimator of \( \beta = (\beta_0, \beta_1)' \) of the model (3.1) is

\[
\hat{\beta}_{IV,1} = \left[ \sum_{i=1}^{n} (z_i - \bar{z}_n)(x_i - \bar{x}_n) \right]^{-1} \sum_{i=1}^{n} (z_i - \bar{z}_n)(y_i - \bar{y}_n),
\]

\[
\hat{\beta}_{IV,0} = \bar{y}_n - \hat{\beta}_{IV,1}\bar{x}_n.
\]

An estimated variance of \( \hat{\beta}_{IV,1} \) is

\[
\hat{V}\{\hat{\beta}_{IV,1}\} = \left[ \sum_{i=1}^{n} (\hat{x}_i - \bar{x}_n)^2 \right]^{-1} \hat{\sigma}^2,
\]

where \( \hat{x}_i \) is the fitted value from the regression of \( x_i \) on \( z_i \), and \( \hat{\sigma}^2 = (n-2)^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_{IV,0} - \hat{\beta}_{IV,1}x_i)^2 \). If \( x_i \) and \( e_i \) are uncorrelated, the asymptotic variance of the IV estimator is always larger, and sometimes much larger, than that of the OLS estimator.

Suppose we have more than one instrumental variable. With multiple instruments, one form of an IV estimator is called the two-stage least squares (2SLS) estimator.
Consider a regression model,

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_{1,i} + e_i, \]  

\[ e_i \sim (0, \sigma^2), \]  

(3.3)

where \( \text{Cov}(x_i, e_i) \neq 0 \) and \( \text{Cov}(z_{1,i}, e_i) = 0 \). Suppose we have two instrumental variables \( z_{2,i} \) and \( z_{3,i} \). Since \( z_{1,i}, z_{2,i} \) and \( z_{3,i} \) are uncorrelated with \( e_i \), any linear combination of \( z_{1,i}, z_{2,i} \) and \( z_{3,i} \) is also uncorrelated with \( e_i \), and therefore any linear combination of \( z_{1,i}, z_{2,i} \) and \( z_{3,i} \) is a valid instrumental variable. We break \( x_i \) into two pieces and let

\[ x_i = x_i^* + v_i, \]  

(3.4)

where \( E(v_i) = 0 \), \( \text{Cov}(z_{1,i}, v_i) = 0 \), \( \text{Cov}(z_{2,i}, v_i) = 0 \) and \( \text{Cov}(z_{3,i}, v_i) = 0 \). The first part is \( x_i^* \) which is uncorrelated with the error term \( e_i \) and \( x_i^* \) is the linear combination that is most highly correlated with \( x_i \). Let

\[ x_i^* = \alpha_0 + \alpha_1 z_{1,i} + \alpha_2 z_{2,i} + \alpha_3 z_{3,i}, \]  

(3.5)

where \( \alpha_2 \neq 0 \) or \( \alpha_3 \neq 0 \). The second piece is \( v_i \) which is possibly correlated with \( e_i \). We obtain the 2SLS estimator in two steps. The first step is to regress \( x_i \) on \( z_{1,i}, z_{2,i} \) and \( z_{3,i} \) and obtain the fitted value \( \hat{x}_i \) which can be used as the estimated best instrumental variable for \( x_i \), where

\[ \hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_{1,i} + \hat{\alpha}_2 z_{2,i} + \hat{\alpha}_3 z_{3,i} \]

is the estimated version of \( x_i^* \). The second step is the ordinary least squares regression \( y_i \) on \( (1, \hat{x}_i, z_{1,i}) \). The 2SLS estimators \( \hat{\beta}_{2SLS,0}, \hat{\beta}_{2SLS,1} \) and \( \hat{\beta}_{2SLS,2} \) are the OLS estimators from the regression \( y_i \) on \( (1, \hat{x}_i, z_{1,i}) \). An estimated variance of \( \hat{\beta}_{2SLS} = (\hat{\beta}_{2SLS,0}, \hat{\beta}_{2SLS,1}, \hat{\beta}_{2SLS,2})' \) is

\[ \hat{V}(\hat{\beta}_{2SLS}) = \begin{pmatrix} n & \sum_{i=1}^{n} \hat{x}_i & \sum_{i=1}^{n} z_{1,i} \\ \sum_{i=1}^{n} \hat{x}_i & \sum_{i=1}^{n} \hat{x}_i^2 & \sum_{i=1}^{n} \hat{x}_i z_{1,i} \\ \sum_{i=1}^{n} z_{1,i} & \sum_{i=1}^{n} \hat{x}_i z_{1,i} & \sum_{i=1}^{n} z_{1,i}^2 \end{pmatrix}^{-1} \sigma^2, \]
\[ \hat{\sigma}^2 = (n - 3)^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_{2SLS,0} - \hat{\beta}_{2SLS,1}\hat{x}_i - \hat{\beta}_{2SLS,2}z_{1,i})^2. \]

The 2SLS estimator is less efficient than the OLS estimator when the explanatory variables are exogenous. Thus it’s helpful to test endogeneity of an explanatory variable to see if the 2SLS estimation procedure is necessary. Testing if \( x_i \) and \( e_i \) are correlated is equivalent to testing if \( v_i \) and \( e_i \) are correlated in the model (3.5) because \( z_{1,i}, z_{2,i} \) and \( z_{3,i} \) are uncorrelated with \( e_i \). The error term \( v_i \) in the model (3.5) is not observed. After fitting model (3.5) by ordinary least squares we obtain the residual \( \hat{v}_i = x_i - \hat{x}_i \). Then we add \( \hat{v}_i \) to the regression equation

\[ y_i = \beta_0 + \beta_1 \hat{x}_i + \beta_2 z_{1,i} + \gamma \hat{v}_i + error. \] (3.6)

If \( x_i \) is exogenous, the coefficient of \( \hat{x}_i \) and the coefficient of \( \hat{v}_i \) are the same, and the null hypothesis is \( H_0 : \beta_1 = \gamma \) for the test of endogeneity. Thus we rewrite the regression model (3.6) as

\[ y_i = \beta_0 + \beta_1 (\hat{x}_i + \hat{v}_i) + \beta_2 z_{1,i} + (\gamma - \beta_1)\hat{v}_i + error \] (3.7)

\[ = \beta_0 + \beta_1 x_i + \beta_2 z_{1,i} + \delta \hat{v}_i + error, \]

and test \( H_0 : \delta = 0 \) using a \( t \) statistic. If the test is used for testing endogeneity of multiple explanatory variables, we use an \( F \) test. If we reject the null hypothesis at some significance level, we conclude that \( x_i \) is endogenous because \( v_i \) and \( e_i \) are correlated. For details see Wooldridge (2000).

To illustrate the instrumental variable approach, Wooldridge (2000) provides an example considering the problem of estimating the causal effect of skipping classes on the final exam score. Consider the regression model

\[ score_i = \beta_0 + \beta_1 skipped_i + e_i, \] (3.8)

where \( score \) is the final exam score, and \( skipped \) is the total number of lectures missed during the semester. The full model might be

\[ score_i = \beta_0 + \beta_1 skipped_i + \beta_2 ability_i + e_i, \] (3.9)
where ability is the student ability. It is difficult to directly measure the student ability and the score is an attempted measurement, therefore the reduced model (3.8) may be proposed. We might be worried that skipped is correlated with other factors in e if we use the model (3.8). Because better students might miss fewer classes. Thus a simple regression model (3.8) of score on skipped may not give us a good estimator for skipped and a good estimate of the causal effect of missing class.

We want a variable that is correlated with skipped, has no direct effect on score, and is not correlated with ability. One option is to use distance between the living place and the campus. Some students will commute to campus, which may increase the likelihood of missing lectures, due to bad weather, traffic, and so on. Thus skipped may be positively correlated with distance.

Next we need to consider if distance is correlated with e in the model (3.8). There are some factors in e may be correlated with distance. For example, student from low-income families may live off campus. If income affects student performance, then distance is correlated with e. Otherwise distance might be a good instrument for skipped.

3.2 Instrumental Variables for Weighted Samples

Assume we have a sample selected from a finite population generated from a superpopulation in which (1.13) holds. Let \((y_i, x_i, r_i, w_i)\) be the vector of observations, where \(w_i = \pi_i^{-1}\). If (1.14) and (1.15) hold for the vector \((x_i, r_i, e_i)\) in the superpopulation, then

\[
E \left\{ \sum_{i \in A} w_i e_i \right\} = E \left\{ \sum_{i \in U} e_i \right\} = 0,
\]

(3.10)

\[
E \left\{ \sum_{i \in A} w_i r_i e_i \right\} = E \left\{ \sum_{i \in U} r_i e_i \right\} = 0,
\]

(3.11)
and

\[
|E \left\{ \sum_{i \in A} w_i x_i' r_i' x_i \right\} | = |E \left\{ \sum_{i \in U} x_i' r_i' x_i \right\} | \neq 0,
\]

(3.12)

Let \( z_i = N^{-1} n w_i r_i \). The \( z_i \) is called the sample instrumental variable or instrument.

If we multiply (1.13) by \( z_i \), and sum we obtain

\[
\sum_{i \in A} z_i y_i = \sum_{i \in A} z_i x_i' \beta + \sum_{i \in A} z_i e_i,
\]

(3.13)

and equation (3.13) in matrix notation is

\[
Z'y = Z'X \beta + \hat{b},
\]

(3.14)

where \( \hat{b} = \sum_{i \in A} z_i e_i \), \( Z \) is an \( n \times k_1 \) matrix, \( X \) is an \( n \times k \) matrix, and \( k_1 \geq k \). Under the model (3.14),

\[
E\{\hat{b}\} = E\{Z'e\} = 0.
\]

If \( e \) is independent of \( Z \), \( V\{Z'e\} = Z'Z \sigma^2_e \). It is reasonable to construct an IV estimator in the form of the 2SLS estimator, where the 2SLS estimator can be written as

\[
\hat{\beta}_{IV} = \left( (Z'X)' (Z'Z)^{-1} Z'X \right)^{-1} (Z'X)' (Z'Z)^{-1} Z'y.
\]

(3.15)

The PW estimator is a special case of the IV estimator (3.15) with \( Z = WX \). By plugging \( Z = WX \) into (3.15), we have

\[
\left[ (Z'X)' (Z'Z)^{-1} Z'X \right]^{-1} (Z'X)' (Z'Z)^{-1} Z'y
= \left[ (X'WX)' (X'WWX)^{-1} (X'WX) \right]^{-1} (X'WX)' (X'WWX)^{-1} (X'Wy)
= [X'WX]^{-1} X'Wy,
\]

provided \( X'WX \) and \( X'WWX \) are nonsingular matrices. Similarly, the PS estimator is a special case of the estimator (3.15) with \( Z = W\hat{W}^{-1} X \), where \( \hat{W} \) is defined in (2.32). The \( z_i = w_i h_i x_i \), where \( h_i \) is defined in (2.1), can be also viewed as an instrumental variable.
3.3 Central Limit Theorem

In this section, we show that the IV estimator is consistent for the population parameter and has a limiting normal distribution.

**Theorem 2.** Let \((y_i, x_i, r_i)\) be a sequence of independent identically distributed random variables with 9-th moment. Let \(\{U_N, \mathcal{F}_N : N = k + 3, k + 4, \ldots\}\) be a sequence of finite populations, where \(U_N\) is the set of indices identifying the elements and \(\mathcal{F}_N = ((y_1, x_1, r_1), \ldots, (y_N, x_N, r_N))\). In the superpopulation \(y_i\) is related to \(x_i\) through the regression model (2.9).

Assume \(r_i\) is independent of \(e_j\) for all \(i\) and \(j\) in the superpopulation and assume that \(E\{(R_N\Gamma_N)'R_N\Gamma_N\}\) is nonsingular, where \(R_N\) is the \(N \times k_1\) matrix of observations on \(r_i\) and \(\Gamma_N = \left\{E(R_N'R_N)\right\}^{-1}E(R_N'X_N)\). Let \(t_j = (y_j, x_j, z_j)\), let

\[
M_{T\pi T,N} = n_N^{-1}T_N'D_{\pi,N}T_N
\]

and

\[
M_{T\pi T} = E\{M_{T\pi T,N}\},
\]

where \(z_i = N^{-1}n_N\pi_i^{-1}r_i\), \(D_{\pi,N} = \text{diag}(\pi_1, \pi_2, \ldots, \pi_N)\), \(\pi_i\) is the inclusion probability for element \(i\), and \(T_N = (t_1', t_2', \ldots, t_N')\). Assume \(K_L < Nn_N^{-1}\pi_i < K_N\) for some positive \(K_L\) and \(K_N\). Assume \(\lim_{N \to \infty} f_N = f\) a.s., where \(f_N = N^{-1}n_N\) and \(0 \leq f < 1\).

Let the condition (2.13) on the Horvitz-Thompson mean in Theorem 1 hold for any \(t\) with 3-rd moments. Let the conditions (2.14) and (2.15) on the Horvitz-Thompson mean in Theorem 1 hold for any \(t\) with 3-rd moments.

Let the instrumental variable estimator be

\[
\hat{\beta}_{IV} = \hat{L}_{XZ}n_N^{-1}Z'y,
\]

where

\[
\hat{L}_{XZ} = [n_N^{-1}X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}.
\]
Then
\[ n_N^{1/2} \left[ V_\infty \{ n_N^{1/2} (\hat{\beta}_{IV} - \beta) \} \right]^{-1/2} [\hat{\beta}_{IV} - \beta] \overset{d}{\to} N(0, I), \tag{3.19} \]
where
\[ V_\infty \{ n_N^{1/2} (\hat{\beta}_{IV} - \beta) \} = L_{XZ} \left[ V_\infty \{ n_N^{1/2} \bar{b} | F_N \} + V \{ n_N^{1/2} \tilde{b}_N \} \right] L_{XZ}', \]
\[ L_{XZ} = E \{ [M_{X\pi Z,N} M_{Z\pi Z,N}^{-1} M_{Z\pi Z,N} M_{Z\pi Z,N}^{-1}] \}, \]
\[ V_\infty \{ n_N^{1/2} \bar{b} | F_N \} \]
is the limiting value of \( V \{ n_N^{1/2} \bar{b} | F_N \} \), \( \bar{b} = n_N^{-1} \sum_{i \in A} b_i \), \( \bar{b}_N = n_N^{-1} \sum_{i \in U} \pi_i b_i \), and \( b_i = z_i e_i \). Then the estimated variance is
\[ \hat{V} \{ \hat{\beta}_{IV} \} = \hat{L}_{XZ} \left[ \hat{V} \{ \tilde{b} \} + n_N^{-2} Z' \hat{D}_e Z \right] \hat{L}_{XZ}' \]
\[ = n_N^{-1} V_\infty \{ n_N^{1/2} (\hat{\beta}_{IV} - \beta) \} + o_p(n_N^{-1}), \tag{3.20} \]
where \( \hat{V} \{ \tilde{b} \} \) is the Horvitz-Thompson estimated variance of \( \tilde{b} \) calculated with \( \tilde{b}_i = z_i \hat{e}_i \), \( \hat{D}_e = \text{diag}(\hat{e}_1^2, \hat{e}_2^2, \ldots, \hat{e}_n^2) \), and \( \hat{e}_i = y_i - x_i' \hat{\beta}_{IV} \).

Proof. By the moment assumptions and assumption (2.13),
\[ (n_N^{-1} \sum_{i \in A} x_i z_i', n_N^{-1} \sum_{i \in A} z_i z_i') - (M_{X\pi Z,N}, M_{Z\pi Z,N}) | F_N = O_p(n_N^{-1/2}) \ a.s., \]
and
\[ (M_{X\pi Z,N}, M_{Z\pi Z,N}) - (M_{X\pi Z}, M_{Z\pi Z}) = O_p(N^{-1/2}). \]
Then
\[ \hat{L}_{XZ} - L_{XZ,N} = O_p(n_N^{-1/2}), \]
where \( L_{XZ,N} = [M_{X\pi Z,N} M_{Z\pi Z,N}^{-1} M_{Z\pi Z,N} M_{Z\pi Z,N}^{-1}]^{-1} M_{X\pi Z,N} M_{Z\pi Z,N}^{-1} \). Similarly
\[ n_N^{-1} \sum_{i \in A} z_i e_i - M_{\pi e,N} | F_N = O_p(n_N^{-1/2}). \]
Under the model (2.9), by assumption, \( e_N \) is independent of \( R_N \). Thus \( M_{\pi e} = 0 \) and
\[ n_N^{-1} \sum_{i \in A} z_i e_i = O_p(n_N^{-1/2}). \]
Thus
\[
\hat{\beta}_{IV} - \beta = \left[ n^{-1}_N X'Z(Z'Z)^{-1}Z'X \right]^{-1} X'Z(Z'Z)^{-1} n^{-1}_N Z'(y - X\beta)
\]
\[
= \hat{L}_{XZ} n^{-1}_N \sum_{i\in A} z_i e_i
\]
\[
= \hat{L}_{XZ} \bar{b}
\]
\[
= L_{XZ, N} \bar{b} + O_p(n^{-1}_N)
\]
\[
= L_{XZ} \bar{b} + O_p(n^{-1}).
\]

Also
\[
\beta_N - \beta = \left[ M_{X\pi Z, N} M_{Z\pi X, N}^{-1} M_{X\pi Z, N} M_{Z\pi X, N}^{-1} n^{-1}_N Z' \right] D_{\pi, N} (y_N - X_N \beta)
\]
\[
= L_{XZ, N} n^{-1}_N \sum_{i\in U} z_i \pi_i e_i
\]
\[
= L_{XZ, N} \bar{b}_N
\]
\[
= L_{XZ} \bar{b}_N + O_p(N^{-1}).
\]

It follows that
\[
\hat{\beta}_{IV} - \beta_N = (\hat{\beta}_{IV} - \beta) - (\beta_N - \beta)
\]
\[
= L_{XZ, N} (\bar{b} - \bar{b}_N) + O_p(n^{-1}_N)
\]
\[
= L_{XZ} (\bar{b} - \bar{b}_N) + O_p(n^{-1}).
\]

From variance assumption (2.13) and the normality assumption (2.14) of the Horvitz-Thompson mean,
\[
n^{-1/2}_N (\bar{b} - \bar{b}_N) | \mathcal{F}_N \overset{\mathcal{L}}{\rightarrow} N(0, V_\infty \{ n^{-1/2}_N \bar{b} | \mathcal{F}_N \}) \quad \text{a.s.,}
\]
where $V_\infty \{ n^{-1/2}_N \bar{b} | \mathcal{F}_N \}$ is defined in (3.19). Then
\[
n^{-1/2}_N (\hat{\beta}_{IV} - \beta_N) | \mathcal{F}_N \overset{\mathcal{L}}{\rightarrow} N(0, L_{XZ} V_\infty \{ n^{-1/2}_N \bar{b} | \mathcal{F}_N \} L'_{XZ}) \quad \text{a.s.} \quad (3.21)
\]

By the Central Limit Theorem for independent random variables and the moment assumptions,
\[
n^{-1/2}_N \bar{b}_N \overset{\mathcal{L}}{\rightarrow} N(0, V \{ n^{-1/2}_N \bar{b} \}).
\]
as $N \to \infty$. Then

$$n_N^{1/2} (\beta_N - \beta) \xrightarrow{L} N(0, L_{XZ} V \{n_N^{1/2} \bar{b}_N \} L'_{XZ}), \quad (3.22)$$

$N \to \infty$.

Applying Lemma 1, result (3.19) then follows from (3.21) and (3.22).

Under the model (2.9),

$$V \{\bar{b}_N\} = V \{N^{-1} \sum \mathbf{r}_i \mathbf{e}_i\} = N^{-2} \sum \mathbf{r}_i \mathbf{r}_i' \sigma^2,$$

thus the estimated variance is

$$\hat{V} \{\bar{b}_N\} = N^{-2} \sum_{i \in A} \pi_i^{-1} \mathbf{r}_i \mathbf{r}_i' \hat{\mathbf{e}}_i^2$$

$$= n_N^{-2} \sum_{i \in A} \pi_i \mathbf{z}_i \mathbf{z}_i' \hat{\mathbf{e}}_i^2$$

$$= n_N^{-2} \mathbf{Z}' \hat{D}_{ae} \hat{D}_{ae} \mathbf{Z},$$

where $\hat{D}_{ae}$ and $\hat{\mathbf{e}}_i$ are defined in (3.20). By the design and the variance assumption (2.15) of the Horvitz-Thompson mean, $\hat{V} \{\bar{b}\} = O_p(n_N^{-1})$, $\hat{V} \{\bar{b}_N\} = O_p(N^{-1})$,

$$\hat{V} \{\bar{b}\} - V\{\bar{b}\} = o_p(n_N^{-1}),$$

and

$$\hat{V} \{\bar{b}_N\} - V\{\bar{b}_N\} = o_p(N^{-1}).$$

Then

$$\hat{L}_{XZ} \left[ \hat{V} \{\bar{b}\} + n_N^{-2} \mathbf{Z}' \hat{D}_{ae} \hat{D}_{ae} \mathbf{Z} \right] \hat{L}'_{XZ} = L_{XZ,N} \left[ \hat{V} \{\bar{b}\} + \hat{V} \{\bar{b}_N\} \right] L'_{XZ,N} + O_p(n_N^{-3/2})$$

$$= L_{XZ,N} \left[ V\{\bar{b}\} \mathcal{F}_N \right] + V\{\bar{b}_N\} \right] L'_{XZ,N} + o_p(n_N^{-1})$$

$$= n_N^{-1} L_{XZ,N} \left[ V\{n_N^{1/2} \bar{b}\} \mathcal{F}_N \right] + V\{n_N^{1/2} \bar{b}_N\} \right]$$

$$L'_{XZ,N} + o_p(n_N^{-1})$$

$$= n_N^{-1} L_{XZ,N} \left[ V\{\infty \{n_N^{1/2} \bar{b}\} \mathcal{F}_N \} + V\{n_N^{1/2} \bar{b}_N\} \right]$$

$$L'_{XZ,N} + o_p(n_N^{-1})$$

$$= n_N^{-1} V\{n_N^{1/2} (\hat{\beta}_IV - \beta)\} + o_p(n_N^{-1}). \quad (3.23)$$
By Theorem 2.2.1 (Fuller, 2006), we can replace \( e_i \) with \( \hat{e}_i \) in (3.23). Result (3.20) then follows.

### 3.4 A Test for Endogeneity

In this section, we describe a test for endogeneity in the context of instrumental variable estimation. Suppose we have a regression model written as

\[
y = X\beta + e, \tag{3.24}
\]

\[
e \sim (0, I\sigma^2).
\]

An \( n \times k_1 \) matrix \( Z \) is a known instrumental variable for \( X \). For example, in the survey situation, a possible \( Z \) is \( Z = WX \). The 2SLS form of the IV estimator constructed using \( Z \) can be written as

\[
\hat{\beta}_{IV} = (\hat{X}'\hat{X})^{-1}\hat{X}'y, \tag{3.25}
\]

where

\[
\hat{X} = Z(Z'Z)^{-1}Z'X.
\]

If the finite population correction can be ignored, an estimated covariance matrix of \( \hat{\beta}_{IV} \) is

\[
\hat{V}\{\hat{\beta}_{IV}\} = (\hat{X}'\hat{X})^{-1}\hat{V}\{\hat{X}'e\}(\hat{X}'\hat{X})^{-1}, \tag{3.26}
\]

where \( \hat{V}\{\hat{X}'e\} \) is a Horvitz-Thompson estimated variance calculated with \( \hat{X}'\hat{e} \) and \( \hat{e}_{i,IV} = y_i - \hat{x}_{i,IV}'\hat{\beta}_{IV} \).

Wooldridge (2000) provides a test for exogeneity based on the 2SLS estimator. We extend the test to the complex survey case. We describe a test of the hypothesis that a set of variables can be used as instrumental variables, given a set that is known to be exogenous. We partition the \( Z \) as \((Z_1, Z_2)\). The \( Z_1 \) is an \( n \times k_{11} \) matrix, the \( Z_2 \) is an \( n \times k_{12} \) matrix and \( k_1 = k_{11} + k_{12} \). The \( Z_1 \) is a set of variables known to be exogenous.
and $Z_2$ is a set for which we wish to test

$$H_0 : E\{Z_2' e\} = 0.$$  \tag{3.27}$$

The test is equivalent to the test that $H_0 : \gamma_2 = 0$ in the representation

$$y = \hat{X} \gamma_1 + (Z_2 - \hat{Z}_2)\gamma_2 + \epsilon,$$  \tag{3.28}$$

where

$$\hat{X} = Z(Z'Z)^{-1}Z'X,$$

and

$$\hat{Z}_2 = Z_1(Z_1'Z_1)^{-1}Z_1'Z_2.$$

We compute

$$\hat{\gamma} = \left(\hat{\gamma}'_1, \hat{\gamma}_2'\right)' = (\hat{X}'\hat{X})^{-1}\hat{X}'y,$$  \tag{3.29}$$

where $\hat{x}_i' = (\hat{x}_i, r_{2i})$ and $r_{2i} = z_{2i} - \hat{z}_{2i}$. If the finite population correction can be ignored, an estimated covariance matrix is

$$\hat{V}\{\hat{\gamma}\} = (\hat{X}'\hat{X})^{-1}\hat{V}\{\hat{X}'\epsilon\}(\hat{X}'\hat{X})^{-1},$$  \tag{3.30}$$

where $\hat{V}\{\hat{X}'\epsilon\}$ is a Horvitz-Thompson estimated variance calculated with $\hat{X}'\hat{\epsilon}$, and $\hat{\epsilon}_i = y_i - \hat{x}_i'\hat{\gamma}$. The null hypothesis is $H_0 : \gamma_2 = 0$ and the test statistic is

$$\hat{\gamma}'_2 \hat{V}_{\gamma\gamma22}^{-1} \hat{\gamma}_2,$$  \tag{3.31}$$

where $k_{12}$ is the dimension of $r_{2i}$, and $\hat{V}_{\gamma\gamma22}$ is the lower right $k_{12} \times k_{12}$ block of $\hat{V}\{\hat{\gamma}\}$. If the test is statistically significant at the chosen significance level, we reject the hypothesis that $Z_2$ can be used as an instrumental variable. Under the null model, the distribution of the test statistic (3.31) is approximately the $\chi^2$ distribution with degrees of freedom equal to $k_{12}$.  

CHAPTER 4  Preliminary Testing Procedure

4.1 Pretest Procedure Example

Because of the frequent use of pretest procedures, we illustrate the properties of the pretest procedure using an example from Huntsberger (1955). Suppose we have a sample \( \{x_1, x_2, \ldots, x_n\} \) drawn from a normal distribution \( N(\mu, \sigma^2) \). For simplicity, let \( \sigma^2 = n \) in order to make the simple sample mean \( \bar{x} \) follow a normal distribution \( N(\mu, 1) \). Suppose the investigator thinks that the true \( \mu \) might be zero. If the data are consistent with \( \mu = 0 \), then \( \mu = 0 \) is used as the estimator. If the data are inconsistent with \( \mu = 0 \), then \( \mu = \bar{x} \) is the estimator. The pretest estimator is based on the \( t \) test

\[
t = \bar{x}[V\{\bar{x}\}]^{-1/2} = \bar{x}.
\]

The pretest estimator is

\[
\hat{\mu} = \begin{cases} 
0 & \text{if } t < t(\alpha) \\
\bar{x} & \text{if } t \geq t(\alpha).
\end{cases}
\]

The mean squared error of the pretest estimator \( \hat{\mu} \) is a function of \( \mu \),

\[
MSE(\hat{\mu}) = \int_{-\infty}^{t(\alpha)} (0 - \mu)^2 f(\bar{x})d\bar{x} + \int_{t(\alpha)}^{\infty} (\bar{x} - \mu)^2 f(\bar{x})d\bar{x}.
\]

Figure 4.1 shows the mean squared error of the pretest estimator \( \hat{\mu} \) versus the true parameter \( \mu \), with the size of the test \( \alpha = 0.05 \) in the one-sided testing procedure. A solid curve is the mean squared error of the pretest estimator \( MSE(\hat{\mu}) \), while the solid straight line at one can be viewed as the mean squared error of the simple sample mean.
\[ MSE(\bar{x}). \] We see that when the true mean \( \mu \) is close to zero, the pretest estimator beats the simple sample mean, and when the true mean is larger than 5, the pretest estimator is almost as good as the simple sample mean. However, when the true mean falls in the interval \([1, 5]\), the simple sample mean has a smaller mean squared error. The procedure is a one-sided testing procedure, but a two-sided testing procedure can be constructed similarly. A simulated experiment for a pretest procedure is conducted by Wu (2004).

4.2 Test for Importance of Weights

If \( E\{x_i\pi_i e_i\} = 0 \), then the OLS estimator is unbiased and the PW estimator is consistent for the superpopulation parameter \( \beta \). The OLS estimator tends to be superior to the PW estimator. If \( E\{x_i\pi_i e_i\} \neq 0 \), then the OLS estimator is biased, but the PW estimator is still consistent for the superpopulation parameter \( \beta \). Therefore, the OLS estimator is preferred under uncorrelated \( x_i\pi_i \) and \( e_i \), and the PW estimator is preferred otherwise. Testing \( E\{x_i\pi_i e_i\} = 0 \) is a test for importance of weights.

The test for importance of weights can be used as a pretest procedure. In our study, the test for importance of weights is used to determine whether weights should be incorporated into the estimation of the parameters. The null hypothesis is

\[
H_0 : E\{((X'WX)^{-1}X'Wy - (X'X)^{-1}X'y)\} = 0. \tag{4.1}
\]

Multiplying (4.1) by \( X'WX \), the hypothesis becomes

\[
H_0 : E\{(WX - X(X'X)^{-1}XX'W)y\} = 0. \tag{4.2}
\]

A test of the hypothesis (4.1) that the expectation of the OLS estimator (1.4) is equal to the expectation of the PW estimator (1.8) can be performed as a test on coefficients in an expanded multiple regression model. We consider to use a standard technique of adding to our basic model the variables for the competing model. The procedure creates
a new variable $Z = WX$ and we can test the hypothesis by testing for the effect of $Z$ on $y$ in the expanded regression model

$$y = X\beta + Z\gamma + e.$$  \hspace{1cm} (4.3)

If the OLS estimator provides an unbiased estimator then $E\{(X'WX)^{-1}X'Wy\} = E\{(X'X)^{-1}X'y\}$, the coefficient for the weighted vector in the model (4.3) will be a zero, that is $\gamma = 0$. The regression coefficient for $Z$ in the the model (4.3) is the regression coefficient for the regression of $(I - X(X'X)^{-1}X')y$ on $(I - X(X'X)^{-1}X')WX$, which is the regression of $e$ on $WX$ after adjusting for $X$. Assuming the regression model (4.3) holds, the ordinary regression test statistic has an $F$ distribution with $k$ and $n - 2k$ degrees of freedom under the $H^*_0: \gamma = 0$, where $k$ is the dimension of $X$. For details see DuMouchel and Duncan (1983) and Fuller (1984).

If the null hypothesis is accepted then one might proceed to fit a regression model of $y$ on $X$ ignoring the weights. If the test indicates that two estimators are estimating different quantities, then it is necessary to incorporate the inclusion probabilities into the estimation procedure.

A pretest procedure based on an importance of weights test is obtained from two regressions: we perform the regression of $y$ on $X$ and $Z$ (full model), and then refit the regression, dropping the $Z$ variables (reduced model). The $F$ statistic

$$F^k_{n-2k} = \frac{(SSE_{red} - SSE_{full})/k}{MSE_{full}}$$  \hspace{1cm} (4.4)

is computed, where $k$ is the dimension of $Z$, $SSE_{full}$ and $SSE_{red}$ are error sum of squares for the full model and the reduced model respectively, and $MSE_{full}$ is mean squared error for the full model. If $F^k_{n-2k}$ is not statistically significant, we use the OLS estimator $\hat{\beta}_{ols}$, otherwise we use the PW estimator $\hat{\beta}_{PW}$. A pretest estimator of $\beta$ is

$$\hat{\beta}_{pre, PW} = \begin{cases} 
\hat{\beta}_{ols} & \text{if } F < F^k_{n-2k}(\alpha) \\
\hat{\beta}_{PW} & \text{if } F \geq F^k_{n-2k}(\alpha), 
\end{cases}$$  \hspace{1cm} (4.5)
where $F_{n-2k}(\alpha)$ is the $1 - \alpha$ quantile of $F$ distribution.

We can also compute an estimated variance for $\hat{\beta}_{pre,PW}$ using the variance estimation procedure appropriate for the estimator (4.5) chosen. Thus

$$
\hat{V}\{\hat{\beta}_{pre,PW}\} = \begin{cases} 
\hat{V}\{\hat{\beta}_{ols}\} & \text{if } F < F_{n-2k}(\alpha) \\
\hat{V}\{\hat{\beta}_{PW}\} & \text{if } F \geq F_{n-2k}(\alpha)
\end{cases}
$$

(4.6)

where $\hat{V}\{\hat{\beta}_{ols}\}$ is defined in (1.5) and $\hat{V}\{\hat{\beta}_{PW}\}$ is defined in (1.12). The estimated variance $\hat{V}\{\hat{\beta}_{pre,PW}\}$ is conditional on the test statistic (4.4), and is not unbiased variance estimator. The estimated variance $\hat{V}\{\hat{\beta}_{pre,PW}\}$ underestimates the variance of $\hat{\beta}_{pre,PW}$. We call the statistic

$$
t_{\beta_{pre,PW}} = [\hat{V}\{\hat{\beta}_{pre,PW}\}]^{-1/2}(\hat{\beta}_{pre,PW} - \beta)
$$

(4.7)

the $t$ statistic for $\hat{\beta}_{pre,PW}$. The distribution of the $t$ statistic is not that of Student’s $t$.

### 4.3 Instrumental Variable Pretest Procedure

The test for endogeneity in the instrumental variable procedure can be used as a pretest procedure. We constructed a general testing procedure for endogeneity in Section 3.4, and we are interested in testing if a second set of variables can be used as instruments, given that we have an initial set that are known to satisfy the requirements for instrumental variables. Consider a superpopulation model

$$
y_i = \beta_0 + \mathbf{x}'_{1,i} \beta_1 + e_i,
$$

(4.8)

where $e_j$ is independent of $\mathbf{x}_i = (1, \mathbf{x}_{1,i})$ for all $i$ and $j$ in the superpopulation. Assume that the selection probabilities have the representation

$$
\pi_i = g_1(\mathbf{x}_i) + g_2(e_i) + u_i,
$$

(4.9)
where $g_1(\cdot)$ and $g_2(\cdot)$ are continuous differentiable functions and $u_i$ is independent of $(x_i, e_i)$. Because under the model (4.8), $e_i$ is independent of $x_i$, then

$$
E \left\{ \sum_{i \in A} (x_{1,i} - \bar{x}_{1,N}) e_i \right\} \\
= E \left\{ \sum_{i \in U} \pi_i (x_{1,i} - \bar{x}_{1,N}) e_i \right\} \\
= E \left\{ \sum_{i \in U} [g_1(x_i) + g_2(e_i) + u_i] (x_{1,i} - \bar{x}_{1,N}) e_i \right\} \\
= E \left\{ \sum_{i \in U} g_1(x_i) (x_{1,i} - \bar{x}_{1,N}) e_i \right\} + E \left\{ \sum_{i \in U} g_2(e_i) (x_{1,i} - \bar{x}_{1,N}) e_i \right\} \\
+ E \left\{ \sum_{i \in U} u_i (x_{1,i} - \bar{x}_{1,N}) e_i \right\} \\
= 0,
$$

(4.10)

where $\bar{x}_{1,N}$ is the population mean for $x_{1,i}$. Also we have

$$
E \left\{ \sum_{i \in A} w_i e_i \right\} = E \left\{ \sum_{i \in U} \pi_i w_i e_i \right\} = E \left\{ \sum_{i \in U} e_i \right\} = 0.
$$

(4.11)

If we multiply (4.8) by $w_i$ and $x_{1,i} - \bar{x}_{1,N}$ and sum we obtain

$$
\sum_{i \in A} w_i y_i = \sum_{i \in A} w_i \beta_0 + \sum_{i \in A} w_i x'_{1,i} \beta_1 + \sum_{i \in A} w_i e_i
$$

(4.12)

and

$$
\sum_{i \in A} (x_{1,i} - \bar{x}_{1,N}) y_i = \sum_{i \in A} (x_{1,i} - \bar{x}_{1,N}) \beta_0 + \sum_{i \in A} (x_{1,i} - \bar{x}_{1,N}) x'_{1,i} \beta_1 + \sum_{i \in A} (x_{1,i} - \bar{x}_{1,N}) e_i,
$$

(4.13)

respectively. Thus

$$
\sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,N})]' y_i = \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,N})]' (1, x'_{1,i}) \beta + \hat{b},
$$

(4.14)

where, by (4.10) and (4.11)

$$
E\{\hat{b}\} = E \left\{ \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,N})]' e_i \right\} = 0.
$$
We obtain the IV estimator for $\beta$ from (4.14)

$$ \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,N})']' y_i = \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,N})']' (1, x_{1,i}') \hat{\beta}_{IV}. \quad (4.15) $$

If $\bar{x}_{1,N}$ is unknown, we can replace $\bar{x}_{1,N}$ with $\bar{x}_{1,\pi}$ in (4.15), and obtain

$$ \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,\pi})']' y_i = \sum_{i \in A} [w_i, (x_{1,i} - \bar{x}_{1,\pi})']' (1, x_{1,i}') \hat{\beta}_{IV}. \quad (4.16) $$

Under assumptions (4.10) and (4.11) and our usual moment assumptions in Theorem 2, the IV estimator of (4.15) is consistent for $\beta$.

In constructing $\hat{\beta}_{IV}$ of (4.16), the vector $(w_i, (x_{1,i} - \bar{x}_{1,\pi})')$ is the instrument for the vector $(1, x_{1,i})$, and the two vectors have the same dimension. Often the dimension of the instrument vector exceeds that of the vector of explanatory variables. For example, $w_i x_{1,i}$, $w_i \bar{w}_i^{-1}$, where $\bar{w}_i$ is defined in (2.32), and $\hat{h}_i w_i$, where $\hat{h}_i$ is estimated $h_i$ in constructing the H estimator, can also be used as instruments. We are interested in testing the set of variables $(x_{1,i} - \bar{x}_{1,\pi})$, given the set $(w_i, w_i x_{1,i})$ which are known to be an instrument for $(1, x_{1,i})$.

A pretest procedure based on a test for endogeneity is obtained from the ordinary least squares fit (3.29) for the model (3.28). The $F$ statistic is defined in (3.31). If $F_{n-k_1}^{k_{12}}$ is not statistically significant, we use $(x_{1,i} - \bar{x}_{1,\pi})$ as an instrument for $(1, x_{1,i})$, otherwise we only use the set $(w_i, w_i x_{1,i})$ to be instrumental variables for $(1, x_{1,i})$. A pretest estimator of $\beta$ is

$$ \hat{\beta}_{pre,IV} = \begin{cases} \hat{\beta}_{IV2} & \text{if } F < F_{n-k_1}^{k_{12}}(\alpha) \\ \hat{\beta}_{IV1} & \text{if } F \geq F_{n-k_1}^{k_{12}}(\alpha), \end{cases} \quad (4.17) $$

where $F_{n-k_1}^{k_{12}}(\alpha)$ is the $1 - \alpha$ quantile of $F$ distribution,

$$ \hat{\beta}_{IV1} = [X'Z_1(Z_1'Z_1)^{-1}Z_1'X]^{-1}X'Z_1(Z_1'Z_1)^{-1}Z_1'y, $$

$$ \hat{\beta}_{IV2} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y, $$
\( x'_i = (1, x'_{1,i}), \ X = (x_1, x_2, \ldots, x_n)' \), \( z'_{1,i} = (w_i, w_i x_{1,i}) \), \( Z_1 = (z_{1,1,} z_{1,2}, \ldots, z_{1,n})' \),

\( z_{2,i} = x_{1,i} - \bar{x}_{1,\pi} \), \( Z_2 = (z_{2,1}, z_{2,2}, \ldots, z_{2,n})' \), \( z'_i = (w_i, w_i x_{1,i}, x_{1,i} - \bar{x}_{1,\pi}) \), and \( Z = (z_1, z_2, \ldots, z_n)' \).
Figure 4.1: Plot of $MSE(\hat{\mu})$ vs. $\mu$
CHAPTER 5  Simulation Design

5.1 Monte Carlo Study Set-up

To illustrate the different estimation procedures and to assess the performance of the estimation procedures, a Monte Carlo simulation study was designed.

We create each sample in the simulation by the following selection procedure. Let \((x_i, e_i, a_i, u_i)\) be a vector, where \(x_i\) is a normal \((0, 0.5)\) random variable, \(e_i\) is a normal \((0, 0.5)\) random variable, \(a_i\) is a normal \((0, 0.5)\) random variable, \(u_i\) is a uniform \((0, 1)\) random variable, and the variables \(x_i, e_i, a_i,\) and \(u_i\) are mutually independent. Let the selection probability \(p_i\) be a function of \(x_i, e_i\) and \(a_i,\)

\[
p_i = p(x_i, e_i, a_i) = ar(x_i) + br(\psi^{0.5}e_i + [1 - \psi]^{0.5}a_i),
\]

where

\[
r(x) = \begin{cases} 
0.025 & \text{if } x < 0.2 \\
0.475(x - 0.20) + 0.025 & \text{if } 0.2 \leq x \leq 1.2 \\
0.5 & \text{if } x > 1.2
\end{cases}
\]

and \(\psi\) is a parameter that is varied in the experiment. The parameter \(\psi\) determines the correlation between \(p_i\) and \(e_i\). The parameter pair \((a, b)\) determines the strength of the correlation between \(p_i\) and \(x_i\) and the strength of the correlation between \(p_i\) and \(e_i\). Let the sum of \(a\) and \(b\) always equal to 2 to ensure that \(p_i \in [0, 1]\). The expectation of \(p_i\) is 0.221.

We use Poisson sampling design. If \(u_i > p_i\), we reject the vector \((x_i, e_i, a_i)\). If \(u_i \leq p_i\),
the vector \((x_i, e_i, a_i)\) is accepted and \(y_i\) is defined by

\[
y_i = 0.5 + x_i + e_i.
\] (5.3)

We draw 1000 selections to create a sample. This procedure gives an expected sample size of about 221. Results are reported for 10000 samples created in this way.

In the experiment, \(\psi\) is varied and four different values of parameter pair \((a, b)\) are chosen. In our study, we constructed three cases. In case one, we set four different parameter pairs, letting \(a_1 = 1\) and \(b_1 = 1\), \(a_2 = 0.75\) and \(b_2 = 1.25\), \(a_3 = 0.5\) and \(b_3 = 1.5\), and \(a_4 = 0.25\) and \(b_4 = 1.75\). We want to check the relationship between the
correlation between $p_i$ and $x_i$ and the efficiency of using the PS estimator. In case two, we let $a = b = 1$ in (5.1) which has the strongest correlation between $p_i$ and $x_i$ and the weakest correlation between $p_i$ and $e_i$ among four pairs of $(a, b)$. In case three, we focus on $a = 0.25$ and $b = 1.75$ in (5.1) which has the weakest correlation between $p_i$ and $x_i$ and the strongest correlation between $p_i$ and $e_i$ among four pairs of $(a, b)$.

5.2 Estimators for the Population Parameter

5.2.1 Pfeffermann-Sverchkov Estimator

In computing the PS estimators, estimated probabilities $\hat{p}_i$'s are constructed. The $q$-weight in the PS estimator defined in (2.32) is $q_i = w_i\hat{p}_i$, where $\hat{p}_i$ is the predicted value from the OLS regression of $p_i$ on $(1, r(x_i))$. An estimated variance matrix is

$$
\hat{V}(\hat{\beta}_{PS}) = (X'QX)^{-1}X'Q\hat{D}_{ee,PS}QX(X'QX)^{-1},
$$

(5.4)

where

$$
\hat{D}_{ee,PS} = diag(\hat{e}_{1,PS}^2, \hat{e}_{2,PS}^2, \ldots, \hat{e}_{n,PS}^2),
$$

$$
\hat{e}_{i,PS} = y_i - x_i'\hat{\beta}_{PS}, \ X = (x_1, x_2, \ldots, x_n)', \ x_i' = (1, x_i).
$$

5.2.2 Instrumental Variable Estimator

Under our regression model, $E\{\sum_{i\in U}e_i\} = 0$ and $E\{\sum_{i\in U}x_ie_i\} = 0$, thus $w_i$ and $w_ix_i$ are instrumental variables for $1$ and $x_i$. We also consider $x_i$ as a potential instrumental variable because of $E\{\sum_{i\in U}\pi_ix_ie_i\} = 0$.

We construct two IV estimators. The first IV estimator is based on four instrumental variables, $w_i, w_ix_i, w_i\hat{p}_i$, and $w_i\hat{p}_ix_i$. The second IV estimator is based on five instrumental variables, $w_i, w_ix_i, w_i\hat{p}_i, w_i\hat{p}_ix_i$, and $x_i$.

The first IV estimator of $\beta$ is

$$
\hat{\beta}_{IV1} = (X'Z_1(Z_1'Z_1)^{-1}Z_1'X)^{-1}X'Z_1(Z_1'Z_1)^{-1}Z_1'y,
$$

(5.5)
where $z_{1,i}' = (w_i, w_i x_i, w_i \hat{p}_i, w_i \hat{p}_i x_i),$ and $Z_1 = (z_{1,1}, z_{1,2}, \ldots, z_{1,n})'$ is an $n \times 4$ matrix. The estimated covariance matrix of $\hat{\beta}_{IV1}$ is

$$\hat{V}(\hat{\beta}_{IV1}) = (\hat{X}' \hat{X})^{-1} \hat{X}' \hat{D}_{ee,IV1} \hat{X} (\hat{X}' \hat{X})^{-1}, \quad (5.6)$$

where

$$\hat{X}_1 = Z_1 (Z_1' Z_1)^{-1} Z_1' X,$$

$$\hat{D}_{ee,IV1} = diag(\hat{e}_{1,IV1}^2, \hat{e}_{2,IV1}^2, \ldots, \hat{e}_{n,IV1}^2),$$

and $\hat{e}_{i,IV1} = y_i - x_i' \hat{\beta}_{IV1}$.

The second IV estimator of $\beta$ is

$$\hat{\beta}_{IV2} = (X' Z (Z' Z)^{-1} Z' X)^{-1} X' Z (Z' Z)^{-1} Z' y, \quad (5.7)$$

where $z_i' = (w_i, w_i x_i, w_i \hat{p}_i, w_i \hat{p}_i x_i, z_{2,i})$, $z_{2,i} = x_i$, $z_2 = (x_1, x_2, \ldots, x_n)'$, and $Z = (Z_1, z_2)$ is an $n \times 5$ matrix. The estimated covariance matrix of $\hat{\beta}_{IV2}$ is

$$\hat{V}(\hat{\beta}_{IV2}) = (\hat{X}' \hat{X})^{-1} \hat{X}' \hat{D}_{ee,IV2} \hat{X} (\hat{X}' \hat{X})^{-1}, \quad (5.8)$$

where

$$\hat{X} = Z (Z' Z)^{-1} Z' X,$$

$$\hat{D}_{ee,IV} = diag(\hat{e}_{1,IV2}^2, \hat{e}_{2,IV2}^2, \ldots, \hat{e}_{n,IV2}^2),$$

and $\hat{e}_{i,IV2} = y_i - x_i' \hat{\beta}_{IV2}$.

The computations of the PS estimation and the IV estimation will be illustrated in Section 5.5.

### 5.2.3 Preliminary Testing Procedures

We construct two simple pretest estimators. One is based on the OLS estimator and the PW estimator and the other one is based on the OLS and the PS estimator.
The pretest based on PW estimator is obtained from two regressions: the regression of \( y_i \) on \((1, x_i, w_i, w_i'x_i)\) (full model) and the regression of \( y_i \) on \((1, x_i)\) (reduced model). The \( F \) statistic

\[
F^2_{n-4} = \frac{(SSE_{red} - SSE_{full})/2}{MSE_{full}} \tag{5.9}
\]

is computed. The pretest estimator of \( \beta \) is

\[
\hat{\beta}_{pre, PW} = \begin{cases} 
\hat{\beta}_{ols} & \text{if } F < F^2_{n-4}(\alpha) \\
\hat{\beta}_{PW} & \text{if } F \geq F^2_{n-4}(\alpha). 
\end{cases} \tag{5.10}
\]

The size of the test \( \alpha = 0.05 \) and \( \alpha = 0.25 \) were used in the simulation. An estimated variance for \( \hat{\beta}_{pre, PW} \) is

\[
\hat{V}\{\hat{\beta}_{pre, PW}\} = \begin{cases} 
\hat{V}\{\hat{\beta}_{ols}\} & \text{if } F < F^2_{n-4}(\alpha) \\
\hat{V}\{\hat{\beta}_{PW}\} & \text{if } F \geq F^2_{n-4}(\alpha), 
\end{cases} \tag{5.11}
\]

where \( \hat{V}\{\hat{\beta}_{ols}\} \) is defined in (1.5) and \( \hat{V}\{\hat{\beta}_{PW}\} \) is defined in (1.12). The \( t \) statistic for \( \hat{\beta}_{pre, PW} \) is given in (4.7).

The pretest based on PS estimator is similarly obtained from two regressions: the regression of \( y_i \) on \((1, x_i, w_i\hat{p}_i, w_i\hat{p}_i'x_i)\) (full model) and the regression of \( y_i \) on \((1, x_i)\) (reduced model). The \( F \) statistic is defined in (5.9). The pretest estimator of \( \beta \) is

\[
\hat{\beta}_{pre, PS} = \begin{cases} 
\hat{\beta}_{ols} & \text{if } F < F^2_{n-4}(\alpha) \\
\hat{\beta}_{PS} & \text{if } F \geq F^2_{n-4}(\alpha). 
\end{cases} \tag{5.12}
\]

An estimated covariance matrix for \( \hat{\beta}_{pre, PS} \) is

\[
\hat{V}\{\hat{\beta}_{pre, PS}\} = \begin{cases} 
\hat{V}\{\hat{\beta}_{ols}\} & \text{if } F < F^2_{n-4}(\alpha) \\
\hat{V}\{\hat{\beta}_{PS}\} & \text{if } F \geq F^2_{n-4}(\alpha), 
\end{cases} \tag{5.13}
\]

where \( \hat{V}\{\hat{\beta}_{ols}\} \) is defined in (1.5) and \( \hat{V}\{\hat{\beta}_{PS}\} \) is defined in (5.4). The \( t \) statistic for \( \hat{\beta}_{pre, PS} \) is

\[
t_{\hat{\beta}_{pre, PS}} = [\hat{V}\{\hat{\beta}_{pre, PS}\}]^{-1/2}(\hat{\beta}_{pre, PS} - \beta). \tag{5.14}
\]
5.2.4 Two-step Preliminary Testing Procedure

We construct a two-step pretest estimator based on the OLS estimator and the two IV estimators. The first step test is a test for importance of weights. The test procedure is same as the simple pretest procedure to construct the pretest estimator \( \hat{\beta}_{pre,PW} \) in Section 5.2.3. If \( F_{n-4}^2 \) in (5.9) is not statistically significant, we use \( \hat{\beta}_{ols} \), otherwise we proceed to the second test.

The second test is a test for endogeneity. We compute the OLS regression of \( y_i \) on \( (\tilde{x}_0, \tilde{x}_i, x_i - \hat{x}_i) \) as defined in (3.29) where \( \tilde{x}_0 \) is the predicted value from the OLS regression of one on \( (w_i, w_i x_i, w_i \hat{p}_i, w_i \hat{p}_i x_i, x_i) \), \( \tilde{x}_i \) is the predicted value from the OLS regression of \( x_i \) on \( (w_i, w_i x_i, w_i \hat{p}_i, w_i \hat{p}_i x_i, x_i) \), and \( \hat{x}_i \) is the predicted value from the OLS regression of \( x_i \) on \( (w_i, w_i x_i, w_i \hat{p}_i, w_i \hat{p}_i x_i) \). The \( t \) statistic for the hypothesis that \( H_0: \delta = 0 \) is

\[
t = \frac{\hat{\delta}}{\hat{v}(\hat{\delta})},
\]

where \( \hat{\delta} \) is the ordinary least squares coefficient for \( x_i - \hat{x}_i \) in the regression of (3.29). Under the null model, the distribution of the \( t \) statistic (5.15) is approximately normal distribution. The two-step pretest estimator is

\[
\hat{\beta}_{pre} = \begin{cases} 
\hat{\beta}_{ols} & \text{if } F < F_{2,n-4}(\alpha) \\
\hat{\beta}_{IV2} & \text{if } |t| < Z(\alpha/2) \\
\hat{\beta}_{IV1} & \text{if } |t| \geq Z(\alpha/2) \text{ and } F \geq F_{2,n-4}(\alpha),
\end{cases}
\]

where \( \alpha \) is the size of the test.

We can compute a standard error for \( \hat{\beta}_{pre} \) using the variance estimation procedure appropriate for the estimator chosen. Then an estimated variance is

\[
\hat{V}\{\hat{\beta}_{pre}\} = \begin{cases} 
\hat{V}\{\hat{\beta}_{ols}\} & \text{if } F < F_{2,n-4}(\alpha) \\
\hat{V}\{\hat{\beta}_{IV2}\} & \text{if } |t| < Z(\alpha/2) \\
\hat{V}\{\hat{\beta}_{IV1}\} & \text{if } |t| \geq Z(\alpha/2) \text{ and } F \geq F_{2,n-4}(\alpha),
\end{cases}
\]
where $\hat{V}\{\hat{\beta}_{IV1}\}$ is defined in (5.6) and $\hat{V}\{\hat{\beta}_{IV2}\}$ is defined in (5.8). The estimated variance $\hat{V}\{\hat{\beta}_{pre}\}$ is conditional on the $F$ test statistic and the $t$ test statistic, and is not unbiased variance estimator. We call the statistic

$$t_{\hat{\beta}_{pre}} = [\hat{V}\{\hat{\beta}_{pre}\}]^{-1/2}(\hat{\beta}_{pre} - \beta)$$

(5.18)

the $t$ statistic for $\hat{\beta}_{pre}$. The distribution of the $t$ statistic is not that of Student’s $t$.

5.3 Simulation Results

5.3.1 Case One Study

In case one, we constructed the OLS estimator, the PW estimator and the PS estimator of $\beta = (\beta_0, \beta_1)$ for four pairs of $(a, b)$. We are interested in knowing that how the performance of the PS estimator depends on the correlation between $p_i$ and $x_i$.

Table 5.1 contains the correlation between $p_i$ and $x_i$ and the correlation between $p_i$ and $e_i$ in the population. The selection probability and $x_i$ are positively correlated. The selection probability and $e_i$ are also positively correlated. For four pairs of $(a, b)$, when $\psi = 0$ in (5.1), $p_i$ and $e_i$ are uncorrelated. If we look at Table 5.1 by column, when $\psi$ increases, the correlation between $p_i$ and $x_i$ is stable, because the correlation between $p_i$ and $x_i$ does not depend on $\psi$ in (5.1). When $\psi$ increases, the correlation between $p_i$ and $e_i$ increases. If we look at Table 5.1 by row, the correlation between $p_i$ and $x_i$ decreases and the correlation between $p_i$ and $e_i$ increases, from left to right, because $a$ decreases and $b$ increases in (5.1).

Table 5.2 contains the means of $p_i x_i$, $p_i x_i^2$ and $p_i e_i$. The means of $p_i x_i$ and $p_i x_i^2$ do not depend on $\psi$. The mean of $p_i e_i$ increases as $\psi$ increases. If we look at Table 5.2 by row, the means of $p_i x_i$ and $p_i x_i^2$ decrease as $a$ decreases, and the decreasing trend for the mean of $p_i x_i$ is faster than that for the mean of $p_i x_i^2$. The mean of $p_i e_i$ increases as
Table 5.1: Monte Carlo Correlations for $corr_1 = corr(p_i, x_i)$ and $corr_2 = corr(p_i, e_i)$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$ $b = 1$</th>
<th>$a = .75$ $b = 1.25$</th>
<th>$a = .5$ $b = 1.5$</th>
<th>$a = .25$ $b = 1.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$corr_1$</td>
<td>$corr_2$</td>
<td>$corr_1$</td>
<td>$corr_2$</td>
</tr>
<tr>
<td>.00</td>
<td>0.572</td>
<td>0.000</td>
<td>0.416</td>
<td>0.000</td>
</tr>
<tr>
<td>.0025</td>
<td>0.572</td>
<td>0.029</td>
<td>0.416</td>
<td>0.035</td>
</tr>
<tr>
<td>.01</td>
<td>0.572</td>
<td>0.057</td>
<td>0.416</td>
<td>0.069</td>
</tr>
<tr>
<td>.02</td>
<td>0.572</td>
<td>0.080</td>
<td>0.416</td>
<td>0.098</td>
</tr>
<tr>
<td>.03</td>
<td>0.572</td>
<td>0.099</td>
<td>0.416</td>
<td>0.120</td>
</tr>
<tr>
<td>.05</td>
<td>0.572</td>
<td>0.128</td>
<td>0.416</td>
<td>0.155</td>
</tr>
<tr>
<td>.07</td>
<td>0.572</td>
<td>0.151</td>
<td>0.416</td>
<td>0.183</td>
</tr>
<tr>
<td>.10</td>
<td>0.572</td>
<td>0.181</td>
<td>0.416</td>
<td>0.219</td>
</tr>
<tr>
<td>.14</td>
<td>0.572</td>
<td>0.214</td>
<td>0.416</td>
<td>0.259</td>
</tr>
<tr>
<td>.17</td>
<td>0.572</td>
<td>0.236</td>
<td>0.416</td>
<td>0.286</td>
</tr>
<tr>
<td>.20</td>
<td>0.572</td>
<td>0.256</td>
<td>0.416</td>
<td>0.310</td>
</tr>
<tr>
<td>.25</td>
<td>0.572</td>
<td>0.285</td>
<td>0.416</td>
<td>0.346</td>
</tr>
<tr>
<td>.30</td>
<td>0.572</td>
<td>0.313</td>
<td>0.416</td>
<td>0.380</td>
</tr>
<tr>
<td>.40</td>
<td>0.572</td>
<td>0.362</td>
<td>0.417</td>
<td>0.439</td>
</tr>
<tr>
<td>.50</td>
<td>0.572</td>
<td>0.404</td>
<td>0.416</td>
<td>0.490</td>
</tr>
</tbody>
</table>

$b$ increases. Because of definitaion of $p_i$ in (5.1), the Monte Carlo means of $p_i x_i e_i$ for all parameter sets in the experiment are always zero.

Table 5.3 contains the bias ratio of $\hat{\beta}_0$, where $BR(\hat{\beta}_0) = [V(\hat{\beta}_0)]^{-1/2}Bias(\hat{\beta}_0)$ is the bias relative to the standard error. The approximate bias for the OLS estimator is

$$E\{\hat{\beta}_{ols} - \beta\} = E\{(X'X)^{-1}X'e\}$$

$$= [E\{N^{-1}X_N' D_{\pi,N} X_N\}]^{-1} E\{N^{-1}X_N' D_{\pi,N} e_N\}$$

$$= |E\{N^{-1}X_N' D_{\pi,N} X_N\}|^{-1} \begin{pmatrix} E\{\pi_i x_i^2\} & -E\{\pi_i x_i\} \\ -E\{\pi_i x_i\} & E\{\pi_i\} \end{pmatrix} \begin{pmatrix} E\{\pi_i e_i\} \\ E\{\pi_i x_i e_i\} \end{pmatrix},$$

where $|E\{N^{-1}X_N' D_{\pi,N} X_N\}|$ is the determinant of the matrix $E\{N^{-1}X_N' D_{\pi,N} X_N\}$. The bias of $\hat{\beta}_0$ principally comes from the product of the positive correlation between $p_i$ and $e_i$ and the positive correlation between $p_i$ and $x_i^2$. If we look at Table 5.3 by column, the bias ratio for $\hat{\beta}_{ols,0}$ increases when $\psi$ increases, because the correlation between $p_i$
Table 5.2: Monte Carlo Means for $E_1 = E\{p_ix_i\}, E_2 = E\{p_ix_i^2\}$ and $E_3 = E\{p_ie_i\}$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$</th>
<th>$b = 1$</th>
<th>$a = 0.75$</th>
<th>$b = 1.25$</th>
<th>$a = 0.5$</th>
<th>$b = 1.5$</th>
<th>$a = 0.25$</th>
<th>$b = 1.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E_1$</td>
<td>$E_2$</td>
<td>$E_3$</td>
<td>$E_1$</td>
<td>$E_2$</td>
<td>$E_3$</td>
<td>$E_1$</td>
<td>$E_2$</td>
</tr>
<tr>
<td>0.00</td>
<td>0.082</td>
<td>0.159</td>
<td>0.000</td>
<td>0.061</td>
<td>0.147</td>
<td>0.000</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.025</td>
<td>0.082</td>
<td>0.159</td>
<td>0.004</td>
<td>0.061</td>
<td>0.147</td>
<td>0.005</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.05</td>
<td>0.082</td>
<td>0.159</td>
<td>0.008</td>
<td>0.061</td>
<td>0.147</td>
<td>0.010</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.075</td>
<td>0.082</td>
<td>0.159</td>
<td>0.012</td>
<td>0.061</td>
<td>0.147</td>
<td>0.014</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.1</td>
<td>0.082</td>
<td>0.159</td>
<td>0.014</td>
<td>0.061</td>
<td>0.147</td>
<td>0.018</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.125</td>
<td>0.082</td>
<td>0.159</td>
<td>0.018</td>
<td>0.061</td>
<td>0.147</td>
<td>0.023</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.15</td>
<td>0.082</td>
<td>0.159</td>
<td>0.022</td>
<td>0.061</td>
<td>0.147</td>
<td>0.027</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.2</td>
<td>0.082</td>
<td>0.159</td>
<td>0.026</td>
<td>0.061</td>
<td>0.147</td>
<td>0.032</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.25</td>
<td>0.082</td>
<td>0.159</td>
<td>0.031</td>
<td>0.061</td>
<td>0.147</td>
<td>0.038</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.3</td>
<td>0.082</td>
<td>0.159</td>
<td>0.034</td>
<td>0.061</td>
<td>0.147</td>
<td>0.042</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.5</td>
<td>0.082</td>
<td>0.159</td>
<td>0.036</td>
<td>0.061</td>
<td>0.147</td>
<td>0.046</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>0.75</td>
<td>0.082</td>
<td>0.159</td>
<td>0.041</td>
<td>0.061</td>
<td>0.147</td>
<td>0.051</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>1.0</td>
<td>0.082</td>
<td>0.159</td>
<td>0.045</td>
<td>0.061</td>
<td>0.147</td>
<td>0.056</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>1.5</td>
<td>0.082</td>
<td>0.159</td>
<td>0.052</td>
<td>0.061</td>
<td>0.147</td>
<td>0.065</td>
<td>0.041</td>
<td>0.135</td>
</tr>
<tr>
<td>2.0</td>
<td>0.082</td>
<td>0.159</td>
<td>0.058</td>
<td>0.061</td>
<td>0.147</td>
<td>0.072</td>
<td>0.041</td>
<td>0.135</td>
</tr>
</tbody>
</table>

and $e_i$ increases. Biases relative to the standard error are close to zero for the PW estimator and the PS estimator, because the PW estimator and the PS estimator are consistent and nonzero biases are due to small sample bias. If we look at Table 5.3 by row, the bias ratio for $\hat{\beta}_{ols,0}$ increases as $b$ increases, that is, the correlation between $p_i$ and $e_i$ increases.

Table 5.4 contains the bias ratio of $\hat{\beta}_1$, where $BR(\hat{\beta}_1) = [V\{\hat{\beta}_1\}]^{-1/2}Bias(\hat{\beta}_1)$. The bias of $\hat{\beta}_1$ principally comes from the product of the positive correlation between $p_i$ and $e_i$ and the positive correlation between $p_i$ and $x_i$. If we look at Table 5.4 by column, the absolute value of the bias ratio for $\hat{\beta}_{ols,1}$ increases when $\psi$ increases, and the bias ratio for $\hat{\beta}_{PW,1}$ and the bias ratio for $\hat{\beta}_{PS,1}$ are nearly zero. If we look at Table 5.4 by row, the absolute value of the bias ratio for $\hat{\beta}_{ols,1}$ decreases as $a$ decreases.

Table 5.5 contains the mean squared errors of $\hat{\beta}_0$. Table 5.6 contains the mean
Table 5.3: Monte Carlo Bias Ratio $\frac{Bias(\hat{\beta}_0)}{\sqrt{\text{V}(\hat{\beta}_0)}}$ for estimators of $\beta_0$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$</th>
<th>$b = 1$</th>
<th>$a = .75$</th>
<th>$b = 1.25$</th>
<th>$a = .5$</th>
<th>$b = 1.5$</th>
<th>$a = .25$</th>
<th>$b = 1.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
<td>$\hat{\beta}_{PS}$</td>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
<td>$\hat{\beta}_{PS}$</td>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
</tr>
<tr>
<td>.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.00</td>
<td>-0.01</td>
</tr>
<tr>
<td>.0025</td>
<td>0.41</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>0.58</td>
<td>0.01</td>
</tr>
<tr>
<td>.01</td>
<td>0.87</td>
<td>0.02</td>
<td>0.02</td>
<td>1.03</td>
<td>0.02</td>
<td>0.02</td>
<td>1.19</td>
<td>0.02</td>
</tr>
<tr>
<td>.02</td>
<td>1.23</td>
<td>0.03</td>
<td>0.03</td>
<td>1.48</td>
<td>0.03</td>
<td>0.03</td>
<td>1.72</td>
<td>0.02</td>
</tr>
<tr>
<td>.03</td>
<td>1.49</td>
<td>0.00</td>
<td>0.00</td>
<td>1.78</td>
<td>0.01</td>
<td>0.01</td>
<td>2.08</td>
<td>0.01</td>
</tr>
<tr>
<td>.05</td>
<td>1.92</td>
<td>0.01</td>
<td>0.01</td>
<td>2.28</td>
<td>0.01</td>
<td>0.01</td>
<td>2.66</td>
<td>0.01</td>
</tr>
<tr>
<td>.07</td>
<td>2.25</td>
<td>0.01</td>
<td>0.01</td>
<td>2.71</td>
<td>0.01</td>
<td>0.01</td>
<td>3.17</td>
<td>0.01</td>
</tr>
<tr>
<td>.10</td>
<td>2.72</td>
<td>0.03</td>
<td>0.03</td>
<td>3.27</td>
<td>0.03</td>
<td>0.03</td>
<td>3.80</td>
<td>0.04</td>
</tr>
<tr>
<td>.14</td>
<td>3.22</td>
<td>0.03</td>
<td>0.03</td>
<td>3.85</td>
<td>0.03</td>
<td>0.03</td>
<td>4.47</td>
<td>0.03</td>
</tr>
<tr>
<td>.17</td>
<td>3.49</td>
<td>0.02</td>
<td>0.02</td>
<td>4.19</td>
<td>0.02</td>
<td>0.02</td>
<td>4.89</td>
<td>0.02</td>
</tr>
<tr>
<td>.20</td>
<td>3.83</td>
<td>0.05</td>
<td>0.05</td>
<td>4.60</td>
<td>0.05</td>
<td>0.05</td>
<td>5.36</td>
<td>0.05</td>
</tr>
<tr>
<td>.25</td>
<td>4.28</td>
<td>0.04</td>
<td>0.04</td>
<td>5.17</td>
<td>0.04</td>
<td>0.04</td>
<td>6.05</td>
<td>0.04</td>
</tr>
<tr>
<td>.30</td>
<td>4.67</td>
<td>0.07</td>
<td>0.07</td>
<td>5.64</td>
<td>0.07</td>
<td>0.07</td>
<td>6.58</td>
<td>0.07</td>
</tr>
<tr>
<td>.40</td>
<td>5.33</td>
<td>0.05</td>
<td>0.05</td>
<td>6.45</td>
<td>0.05</td>
<td>0.05</td>
<td>7.58</td>
<td>0.05</td>
</tr>
<tr>
<td>.50</td>
<td>6.03</td>
<td>0.06</td>
<td>0.06</td>
<td>7.29</td>
<td>0.06</td>
<td>0.06</td>
<td>8.54</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Squared errors of $\hat{\beta}_1$. The mean squared errors of $\hat{\beta}_{ols,0}$ and $\hat{\beta}_{ols,1}$ are the smallest among estimators of $\beta_0$ and $\beta_1$ in every pair of $(a, b)$, when $\psi = 0$, that is, when there is no correlation between $p_i$ and $e_i$. If we look at Table 5.5 by column, we see that increasing the correlation between $p_i$ and $e_i$ increases the mean squared errors of $\hat{\beta}_{ols,0}$ because of the squared bias. The mean squared errors of the PW estimators and the mean squared errors of the PS estimators have slightly decreasing trend for every pair of $(a, b)$, when $\psi$ increases, because of increasing correlation between $\pi_i$ and $e_i$.

When $a = b = 1$, the estimators $\hat{\beta}_{PS,0}$ are more efficient than $\hat{\beta}_{PW,0}$, because the correlation between selection probabilities and $x_i$ is relatively high. From left to right, $a$ gets smaller and $b$ gets larger, and the efficiency of the PS estimator is decreasing. When $a = 0.25$ and $b = 1.75$, the PW estimator is almost as good as the PS estimator. The efficiency of the PS estimator reduces, as the parameter $a$ decreases, due to the
Table 5.4: Monte Carlo Bias Ratio $\frac{Bias(\hat{\beta})}{\sqrt{|V(\hat{\beta})|}}$ for estimators of $\beta_1$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$</th>
<th>$b = 1$</th>
<th>$a = .75$</th>
<th>$b = 1.25$</th>
<th>$a = .5$</th>
<th>$b = 1.5$</th>
<th>$a = .25$</th>
<th>$b = 1.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
<td>$\hat{\beta}_{PS}$</td>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
<td>$\hat{\beta}_{PS}$</td>
<td>$\hat{\beta}_{ols}$</td>
<td>$\hat{\beta}_{PW}$</td>
<td>$\hat{\beta}_{PS}$</td>
</tr>
<tr>
<td>.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>.0025</td>
<td>-0.18</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>.01</td>
<td>-0.36</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.36</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.29</td>
<td>-0.01</td>
</tr>
<tr>
<td>.02</td>
<td>-0.51</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.51</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.41</td>
<td>-0.03</td>
</tr>
<tr>
<td>.03</td>
<td>-0.49</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.49</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.25</td>
<td>-0.03</td>
</tr>
<tr>
<td>.05</td>
<td>-0.62</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.62</td>
<td>-0.02</td>
<td>-0.01</td>
<td>-0.38</td>
<td>-0.02</td>
</tr>
<tr>
<td>.07</td>
<td>-0.74</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.74</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.44</td>
<td>0.01</td>
</tr>
<tr>
<td>.10</td>
<td>-0.90</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.90</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.53</td>
<td>-0.02</td>
</tr>
<tr>
<td>.14</td>
<td>-1.06</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-1.06</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.64</td>
<td>-0.01</td>
</tr>
<tr>
<td>.17</td>
<td>-1.15</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-1.15</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.69</td>
<td>-0.01</td>
</tr>
<tr>
<td>.20</td>
<td>-1.25</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-1.25</td>
<td>-0.02</td>
<td>-0.02</td>
<td>-0.75</td>
<td>-0.02</td>
</tr>
<tr>
<td>.25</td>
<td>-1.43</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-1.43</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-0.86</td>
<td>-0.03</td>
</tr>
<tr>
<td>.30</td>
<td>-1.53</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-1.53</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.93</td>
<td>-0.04</td>
</tr>
<tr>
<td>.40</td>
<td>-1.77</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-1.77</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-1.07</td>
<td>-0.04</td>
</tr>
<tr>
<td>.50</td>
<td>-1.97</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-1.97</td>
<td>-0.05</td>
<td>-0.04</td>
<td>-1.20</td>
<td>-0.03</td>
</tr>
</tbody>
</table>

decreasing correlation between $p_i$ and $x_i$. Table 5.6 shows the same tendency as Table 5.5.

Table 5.7 contains the mean squared error ratio of $\hat{\beta}_{PS,0}$ relative to that of $\hat{\beta}_{PW,0}$ for four pairs of $(a,b)$. Table 5.8 contains the mean squared error ratio of $\hat{\beta}_{PS,1}$ relative to that of $\hat{\beta}_{PW,1}$ for four pairs of $(a,b)$. If we look at Table 5.7 by column, each column contains the values of $MSE(\hat{\beta}_{PW,0})/MSE(\hat{\beta}_{PS,0})$ for each pair of $(a,b)$. The MSE ratio values for different $\psi$ are stable in each column. If we look at Table 5.7 by row, the MSE ratios get smaller as $a$ decreases. When the correlation between $p_i$ and $x_i$ is fairly high, the PS estimator works better than the PW estimator. The mean squared error ratio in the last column in Table 5.7 is near one, even though there exists a weak correlation between $p_i$ and $x_i$. Because of the variance of the estimated $p_i$ in the PS estimation increases the total variance of the PS estimator. Table 5.8 shows the same tendency as
Table 5.5: Monte Carlo Mean Squared Error (×1000) for estimators of β₀ (10,000 samples)

<table>
<thead>
<tr>
<th>ψ</th>
<th>a = 1 b = 1</th>
<th>a = .75 b = 1.25</th>
<th>a = .5 b = 1.5</th>
<th>a = .25 b = 1.75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β₀₀</td>
<td>β₀₁</td>
<td>β₀₀</td>
<td>β₀₁</td>
</tr>
<tr>
<td>.00</td>
<td>2.86</td>
<td>5.38</td>
<td>5.13</td>
<td>2.60</td>
</tr>
<tr>
<td>.01</td>
<td>4.96</td>
<td>5.30</td>
<td>5.08</td>
<td>5.35</td>
</tr>
<tr>
<td>.03</td>
<td>8.98</td>
<td>5.34</td>
<td>5.13</td>
<td>10.67</td>
</tr>
<tr>
<td>.07</td>
<td>17.45</td>
<td>5.30</td>
<td>5.08</td>
<td>21.77</td>
</tr>
<tr>
<td>.14</td>
<td>31.93</td>
<td>5.07</td>
<td>4.85</td>
<td>40.81</td>
</tr>
<tr>
<td>.20</td>
<td>44.54</td>
<td>5.10</td>
<td>4.88</td>
<td>57.23</td>
</tr>
<tr>
<td>.30</td>
<td>65.23</td>
<td>5.06</td>
<td>4.85</td>
<td>84.38</td>
</tr>
<tr>
<td>.50</td>
<td>107.22</td>
<td>4.84</td>
<td>4.64</td>
<td>138.93</td>
</tr>
</tbody>
</table>

Table 5.7.

5.3.2 Case Two Study

In the case two, we constructed the OLS estimator, the PW estimator, the PS estimator, and two pretest estimators of \( \beta = (\beta₀, \beta₁) \) for \( a = b = 1 \). The two pretest estimators are defined in Section 5.2.3. We want to assess the performance of the pretest estimators and compare the pretest estimators with other estimators.

Table 5.9 contains the mean squared errors of \( \hat{\beta}_₀ \). Table 5.10 contains the mean squared errors of \( \hat{\beta}_₁ \). The pretest estimators are for \( \alpha = 0.05 \) and \( \alpha = 0.25 \). The pretest estimators based on the PW estimators \( \hat{\beta}_{pre,PW,0} \) and \( \hat{\beta}_{pre,PW,1} \) are uniformly inferior to the pretest estimators based on the PS estimators \( \hat{\beta}_{pre,PS,0} \) and \( \hat{\beta}_{pre,PS,0} \) in terms of mean squared error for both \( \alpha = 0.05 \) and \( \alpha = 0.25 \). When \( \psi \) gets larger, the mean
Table 5.6: Monte Carlo Mean Squared Error ($\times 1000$) for estimators of $\beta_1$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$ $b = 1$</th>
<th>$a = 0.75$ $b = 1.25$</th>
<th>$a = 0.5$ $b = 1.5$</th>
<th>$a = 0.25$ $b = 1.75$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_{ols}$</td>
<td>$\beta_{PW}$</td>
<td>$\beta_{PS}$</td>
<td>$\beta_{pre,PW}$</td>
</tr>
<tr>
<td>0.00</td>
<td>3.99 8.43 6.80</td>
<td>3.90 8.53 7.54</td>
<td>3.95 8.87 8.49</td>
<td>4.19 9.81 9.84</td>
</tr>
<tr>
<td>0.0025</td>
<td>4.09 8.47 6.76</td>
<td>4.02 8.57 7.53</td>
<td>4.05 8.87 8.47</td>
<td>4.26 9.77 9.79</td>
</tr>
<tr>
<td>0.01</td>
<td>4.55 8.35 6.68</td>
<td>4.46 8.43 7.42</td>
<td>4.37 8.86 8.46</td>
<td>4.39 9.81 9.82</td>
</tr>
<tr>
<td>0.02</td>
<td>5.13 8.54 6.83</td>
<td>4.93 8.62 7.59</td>
<td>4.65 8.94 8.55</td>
<td>4.45 9.79 9.82</td>
</tr>
<tr>
<td>0.03</td>
<td>5.57 8.40 6.73</td>
<td>5.35 8.52 7.52</td>
<td>5.03 8.88 8.50</td>
<td>4.65 9.83 9.85</td>
</tr>
<tr>
<td>0.05</td>
<td>6.68 8.38 6.68</td>
<td>6.30 8.51 7.47</td>
<td>5.59 8.89 8.50</td>
<td>4.79 9.80 9.82</td>
</tr>
<tr>
<td>0.07</td>
<td>7.84 8.48 6.78</td>
<td>7.23 8.57 7.54</td>
<td>6.17 8.85 8.45</td>
<td>4.96 9.65 9.67</td>
</tr>
<tr>
<td>0.10</td>
<td>9.50 8.30 6.64</td>
<td>8.77 8.39 7.39</td>
<td>7.28 8.68 8.30</td>
<td>5.46 9.47 9.50</td>
</tr>
<tr>
<td>0.14</td>
<td>11.75 8.41 6.69</td>
<td>10.64 8.50 7.47</td>
<td>8.46 8.78 8.39</td>
<td>5.87 9.64 9.66</td>
</tr>
<tr>
<td>0.20</td>
<td>14.81 8.00 6.42</td>
<td>13.23 8.13 7.17</td>
<td>10.21 8.43 8.08</td>
<td>6.44 9.33 9.36</td>
</tr>
<tr>
<td>0.25</td>
<td>17.81 8.12 6.53</td>
<td>15.92 8.26 7.29</td>
<td>12.06 8.61 8.24</td>
<td>7.17 9.49 9.51</td>
</tr>
<tr>
<td>0.30</td>
<td>20.87 7.96 6.36</td>
<td>18.42 8.05 7.08</td>
<td>13.60 8.34 7.97</td>
<td>7.70 9.17 9.19</td>
</tr>
<tr>
<td>0.40</td>
<td>26.45 7.92 6.33</td>
<td>23.35 8.00 7.05</td>
<td>16.96 8.28 7.93</td>
<td>9.00 8.99 9.03</td>
</tr>
<tr>
<td>0.50</td>
<td>32.09 7.83 6.30</td>
<td>28.19 7.92 7.00</td>
<td>20.24 8.15 7.81</td>
<td>10.44 8.97 9.01</td>
</tr>
</tbody>
</table>

SQUARED ERRORS OF $\hat{\beta}_{pre,PW,0}$ AND $\hat{\beta}_{pre,PW,1}$ ARE CLOSER TO THE MEAN SQUARED ERRORS OF $\hat{\beta}_{PW,0}$ AND $\hat{\beta}_{PW,1}$, RESPECTIVELY, FOR BOTH $\alpha = 0.05$ AND $\alpha = 0.25$. THE REASON FOR THIS TREAT IS THAT THE PRETEST PROCEDURE REJECTS THE NULL HYPOTHESIS MORE FREQUENTLY WHEN THE CORRELATION BETWEEN $p_i$ AND $e_i$ INCREASES. THE TENDENCY FOR THE MEAN SQUARED ERRORS OF THE PRETEST ESTIMATORS BASED ON THE PW ESTIMATORS GETTING CLOSER TO THE MEAN SQUARED ERRORS OF THE PW ESTIMATORS IS MORE OBIOS WITH $\alpha = 0.25$ THAN WITH $\alpha = 0.05$. THE REASON FOR THIS DIFFERENCE IS THAT THE PRETEST PROCEDURE WITH LARGER SIZE OF TEST REJECTS THE NULL HYPOTHESIS MORE FREQUENTLY THAN THE PRETEST PROCEDURE WITH RELATIVELY SMALLER SIZE OF TEST. THE PS ESTIMATORS AND THE PRETEST ESTIMATORS BASED ON THE PS ESTIMATORS SHOW THE SAME TENDENCY AS THE PW ESTIMATORS AND THE PRETEST ESTIMATORS BASED ON THE PW ESTIMATORS.

AS THE SIMULATION RESULTS OF TABLE 5.11 ILLUSTRATE, FOR $\alpha = 0.05$ AND $\alpha = 0.25$, ALMOST ALL STATISTICS $t_{\hat{\beta}_{ols,0}}$, $t_{\hat{\beta}_{PW,0}}$, $t_{\hat{\beta}_{PS,0}}$, $t_{\hat{\beta}_{pre,PW,0}}$ AND $t_{\hat{\beta}_{pre,PS,0}}$ EXCEED THE TABULAR $t_{0.025}$ FOR STUDENT’S
Table 5.7: Monte Carlo Mean Squared Error Ratio $\frac{MSE(\hat{\beta}_{PW,0})}{MSE(\hat{\beta}_{PS,0})}$ for estimators of $\beta_0$

(10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$</th>
<th>$a = .75$</th>
<th>$a = .5$</th>
<th>$a = .25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b = 1$</td>
<td>$b = 1.25$</td>
<td>$b = 1.5$</td>
<td>$b = 1.75$</td>
</tr>
<tr>
<td>.00</td>
<td>1.049</td>
<td>1.027</td>
<td>1.010</td>
<td>0.999</td>
</tr>
<tr>
<td>.0025</td>
<td>1.045</td>
<td>1.025</td>
<td>1.009</td>
<td>1.000</td>
</tr>
<tr>
<td>.01</td>
<td>1.043</td>
<td>1.025</td>
<td>1.009</td>
<td>1.000</td>
</tr>
<tr>
<td>.02</td>
<td>1.047</td>
<td>1.026</td>
<td>1.010</td>
<td>1.000</td>
</tr>
<tr>
<td>.03</td>
<td>1.040</td>
<td>1.023</td>
<td>1.008</td>
<td>0.999</td>
</tr>
<tr>
<td>.05</td>
<td>1.044</td>
<td>1.025</td>
<td>1.010</td>
<td>1.000</td>
</tr>
<tr>
<td>.07</td>
<td>1.044</td>
<td>1.025</td>
<td>1.009</td>
<td>1.000</td>
</tr>
<tr>
<td>.10</td>
<td>1.051</td>
<td>1.028</td>
<td>1.010</td>
<td>1.000</td>
</tr>
<tr>
<td>.14</td>
<td>1.046</td>
<td>1.026</td>
<td>1.011</td>
<td>1.001</td>
</tr>
<tr>
<td>.17</td>
<td>1.044</td>
<td>1.025</td>
<td>1.010</td>
<td>1.000</td>
</tr>
<tr>
<td>.20</td>
<td>1.044</td>
<td>1.025</td>
<td>1.009</td>
<td>1.000</td>
</tr>
<tr>
<td>.25</td>
<td>1.051</td>
<td>1.029</td>
<td>1.010</td>
<td>0.999</td>
</tr>
<tr>
<td>.30</td>
<td>1.044</td>
<td>1.025</td>
<td>1.009</td>
<td>1.000</td>
</tr>
<tr>
<td>.40</td>
<td>1.048</td>
<td>1.027</td>
<td>1.010</td>
<td>0.999</td>
</tr>
<tr>
<td>.50</td>
<td>1.044</td>
<td>1.025</td>
<td>1.009</td>
<td>0.999</td>
</tr>
</tbody>
</table>

$t$ by more than the nominal fraction. As $\psi$ increases, the probabilities of $P(|t_{\hat{\beta}_{PW,0}}| > t_{.025})$ are closer to the probabilities of $P(|t_{\hat{\beta}_{PS,0}}| > t_{.025})$. The $P(|t_{\hat{\beta}_{pre,PS,0}}| > t_{.025})$ and $P(|t_{\hat{\beta}_{PS,0}}| > t_{.025})$ show the same trend. Table 5.12 gives the probabilities of the statistics $t_{\hat{\beta}_{ds,1}}$, $t_{\hat{\beta}_{PW,1}}$, $t_{\hat{\beta}_{PS,1}}$, $t_{\hat{\beta}_{pre,PS,1}}$ and $t_{\hat{\beta}_{pre,PS,1}}$ exceeding the tabular $t_{.025}$. We can see the same tendency in Table 5.12 as in Table 5.11.

Figure 5.2 is the plot of mean squared error ratios of $\hat{\beta}_{ds,0}$, $\hat{\beta}_{PW,0}$ and $\hat{\beta}_{pre,PS,0}$ relative to $\hat{\beta}_{PS,0}$ as a function of correlation between $p_i$ and $e_i$ for $\alpha = 0.05$ and $\alpha = 0.25$. The shapes are typical of preliminary testing procedures. In Figure 5.2 the horizontal line always equal to one can be viewed as the mean squared error efficiency of $\hat{\beta}_{PS,0}$ relative to itself, or the mean squared error efficiency of $\hat{\beta}_{pre,PS,0}$ relative to $\hat{\beta}_{PS,0}$ when $\alpha = 1$. When $\alpha = 1$ we always reject $\hat{\beta}_{ds,0}$ in the pretest procedures. Since $0.05 < 0.25 < 1$, the curve for the mean squared error efficiency of $\hat{\beta}_{pre,PS,0}$ relative to $\hat{\beta}_{PS,0}$ with $\alpha = 0.25$ is
Table 5.8: Monte Carlo Mean Squared Error Ratio $\frac{MSE(\hat{\beta}_{PW,1})}{MSE(\hat{\beta}_{PS,1})}$ for estimators of $\beta_1$
(10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$a = 1$</th>
<th>$a = .75$</th>
<th>$a = .5$</th>
<th>$a = .25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b = 1$</td>
<td>$b = 1.25$</td>
<td>$b = 1.5$</td>
<td>$b = 1.75$</td>
</tr>
<tr>
<td>.00</td>
<td>1.239</td>
<td>1.132</td>
<td>1.045</td>
<td>0.997</td>
</tr>
<tr>
<td>.0025</td>
<td>1.253</td>
<td>1.138</td>
<td>1.046</td>
<td>0.998</td>
</tr>
<tr>
<td>.01</td>
<td>1.249</td>
<td>1.137</td>
<td>1.047</td>
<td>0.998</td>
</tr>
<tr>
<td>.02</td>
<td>1.251</td>
<td>1.137</td>
<td>1.046</td>
<td>0.997</td>
</tr>
<tr>
<td>.03</td>
<td>1.247</td>
<td>1.134</td>
<td>1.045</td>
<td>0.998</td>
</tr>
<tr>
<td>.05</td>
<td>1.256</td>
<td>1.138</td>
<td>1.046</td>
<td>0.997</td>
</tr>
<tr>
<td>.07</td>
<td>1.251</td>
<td>1.137</td>
<td>1.047</td>
<td>0.998</td>
</tr>
<tr>
<td>.10</td>
<td>1.251</td>
<td>1.135</td>
<td>1.046</td>
<td>0.997</td>
</tr>
<tr>
<td>.14</td>
<td>1.258</td>
<td>1.138</td>
<td>1.047</td>
<td>0.997</td>
</tr>
<tr>
<td>.17</td>
<td>1.254</td>
<td>1.137</td>
<td>1.045</td>
<td>0.996</td>
</tr>
<tr>
<td>.20</td>
<td>1.246</td>
<td>1.133</td>
<td>1.043</td>
<td>0.997</td>
</tr>
<tr>
<td>.25</td>
<td>1.243</td>
<td>1.134</td>
<td>1.045</td>
<td>0.998</td>
</tr>
<tr>
<td>.30</td>
<td>1.251</td>
<td>1.137</td>
<td>1.046</td>
<td>0.998</td>
</tr>
<tr>
<td>.40</td>
<td>1.251</td>
<td>1.135</td>
<td>1.045</td>
<td>0.996</td>
</tr>
<tr>
<td>.50</td>
<td>1.243</td>
<td>1.132</td>
<td>1.043</td>
<td>0.995</td>
</tr>
</tbody>
</table>

flatter than with $\alpha = 0.05$. The curve for the mean squared error efficiency of $\hat{\beta}_{pre,PS,0}$ relative to $\hat{\beta}_{PS,0}$ with $\alpha = 0.25$ is generally between the curve of mean squared error efficiency of $\hat{\beta}_{pre,PS,0}$ relative to $\hat{\beta}_{PS,0}$ with $\alpha = 0.05$ and the horizontal line at one.

The solid curve is the mean squared error ratio of $\hat{\beta}_{ols,0}$ relative to $\hat{\beta}_{PS,0}$ as a function of correlation between $p_i$ and $e_i$. The $\hat{\beta}_{ols,0}$ is the best if $p_i$ and $e_i$ are independent, but has very poor performance when the correlation between $p_i$ and $e_i$ is large. The dashed curve for the mean squared error ratio of $\hat{\beta}_{PW,0}$ relative to $\hat{\beta}_{PS,0}$ is always above the horizontal line at one. The $\hat{\beta}_{PW,0}$ is less efficient than the $\hat{\beta}_{PS,0}$ uniformly. The pretest estimator $\hat{\beta}_{pre,PS,0}$ is better than $\hat{\beta}_{PS,0}$, but worse than $\hat{\beta}_{ols,0}$ when the correlation between $p_i$ and $e_i$ is low. When the correlation gets larger, $\hat{\beta}_{pre,PS,0}$ is worse than $\hat{\beta}_{PS,0}$, but better than $\hat{\beta}_{ols,0}$. The pretest estimator $\hat{\beta}_{pre,PS,0}$ is never the best, nor the worst, so the pretest estimator is a compromise in terms of mean squared error.
Table 5.9: Monte Carlo Mean Squared Error (×1000) for estimators of $\beta_0$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{pre,PW,\alpha = 0.05}$</th>
<th>$\hat{\beta}_{pre,PS,\alpha = 0.05}$</th>
<th>$\hat{\beta}_{pre,PW,\alpha = 0.25}$</th>
<th>$\hat{\beta}_{pre,PS,\alpha = 0.25}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>2.76</td>
<td>5.31</td>
<td>5.05</td>
<td>3.26</td>
<td>3.21</td>
<td>4.28</td>
<td>4.11</td>
</tr>
<tr>
<td>.0025</td>
<td>3.40</td>
<td>5.34</td>
<td>5.09</td>
<td>3.98</td>
<td>3.92</td>
<td>4.72</td>
<td>4.58</td>
</tr>
<tr>
<td>.01</td>
<td>4.85</td>
<td>5.28</td>
<td>5.04</td>
<td>5.42</td>
<td>5.35</td>
<td>5.46</td>
<td>5.30</td>
</tr>
<tr>
<td>.02</td>
<td>7.05</td>
<td>5.40</td>
<td>5.16</td>
<td>7.26</td>
<td>7.00</td>
<td>6.39</td>
<td>6.05</td>
</tr>
<tr>
<td>.03</td>
<td>8.97</td>
<td>5.25</td>
<td>5.00</td>
<td>8.42</td>
<td>7.85</td>
<td>6.71</td>
<td>6.19</td>
</tr>
<tr>
<td>.05</td>
<td>13.24</td>
<td>5.28</td>
<td>5.03</td>
<td>9.94</td>
<td>8.65</td>
<td>6.92</td>
<td>6.03</td>
</tr>
<tr>
<td>.07</td>
<td>17.33</td>
<td>5.30</td>
<td>5.06</td>
<td>10.20</td>
<td>8.19</td>
<td>6.63</td>
<td>5.72</td>
</tr>
<tr>
<td>.10</td>
<td>23.59</td>
<td>5.31</td>
<td>5.07</td>
<td>9.38</td>
<td>7.08</td>
<td>6.11</td>
<td>5.34</td>
</tr>
<tr>
<td>.14</td>
<td>32.05</td>
<td>5.13</td>
<td>4.91</td>
<td>7.66</td>
<td>5.64</td>
<td>5.44</td>
<td>4.97</td>
</tr>
<tr>
<td>.17</td>
<td>38.52</td>
<td>5.18</td>
<td>4.93</td>
<td>6.68</td>
<td>5.27</td>
<td>5.33</td>
<td>4.95</td>
</tr>
<tr>
<td>.20</td>
<td>44.51</td>
<td>5.16</td>
<td>4.93</td>
<td>5.98</td>
<td>5.05</td>
<td>5.21</td>
<td>4.94</td>
</tr>
<tr>
<td>.25</td>
<td>54.51</td>
<td>5.04</td>
<td>4.81</td>
<td>5.29</td>
<td>4.83</td>
<td>5.05</td>
<td>4.81</td>
</tr>
<tr>
<td>.30</td>
<td>65.61</td>
<td>5.11</td>
<td>4.88</td>
<td>5.19</td>
<td>4.88</td>
<td>5.12</td>
<td>4.88</td>
</tr>
<tr>
<td>.40</td>
<td>85.76</td>
<td>4.76</td>
<td>4.56</td>
<td>4.76</td>
<td>4.56</td>
<td>4.76</td>
<td>4.56</td>
</tr>
<tr>
<td>.50</td>
<td>107.01</td>
<td>4.90</td>
<td>4.68</td>
<td>4.90</td>
<td>4.68</td>
<td>4.90</td>
<td>4.68</td>
</tr>
</tbody>
</table>

Figure 5.3 is the plot of the mean squared error ratios of $\hat{\beta}_{ols,1}$, $\hat{\beta}_{PW,1}$ and $\hat{\beta}_{pre,PS,1}$ relative to $\hat{\beta}_{PS,1}$ for $\alpha = 0.05$ and $\alpha = 0.25$. The curves for the mean squared error efficiency relative to $\hat{\beta}_{PS,1}$ in Figure 5.3 are similar to the curves of the mean squared error efficiency relative to $\hat{\beta}_{PS,0}$ in Figure 5.2. The two curves for the pretest estimators in Figure 5.3 are smoother than the two pretest curves in Figure 5.2. One reason for this difference is that the bias of $\hat{\beta}_{pre,PS,0}$ relative to the standard error is larger than that of $\hat{\beta}_{pre,PS,1}$.

5.3.3 Case Three Study

In the case three, we constructed the OLS estimator, the PW estimator, the PS estimator, two IV estimators and a two-step pretest estimator of $\beta = (\beta_0, \beta_1)$ when $a = 0.25$ and $b = 1.75$. The two-step pretest estimator is defined in Section 5.2.4. We want to assess the performance of the two IV estimators and the two-step pretest
Table 5.10: Monte Carlo Mean Squared Error (×1000) for estimators of $\beta_1$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{pre, PW}$ $\alpha = 0.05$</th>
<th>$\hat{\beta}_{pre, PS}$ $\alpha = 0.05$</th>
<th>$\hat{\beta}_{pre, PW}$ $\alpha = 0.25$</th>
<th>$\hat{\beta}_{pre, PS}$ $\alpha = 0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>3.99</td>
<td>8.56</td>
<td>6.85</td>
<td>4.90</td>
<td>4.56</td>
<td>6.67</td>
<td>5.68</td>
</tr>
<tr>
<td>.0025</td>
<td>4.06</td>
<td>8.40</td>
<td>6.75</td>
<td>4.97</td>
<td>4.68</td>
<td>6.74</td>
<td>5.75</td>
</tr>
<tr>
<td>.01</td>
<td>4.45</td>
<td>8.47</td>
<td>6.77</td>
<td>5.76</td>
<td>5.37</td>
<td>7.37</td>
<td>6.25</td>
</tr>
<tr>
<td>.02</td>
<td>5.08</td>
<td>8.48</td>
<td>6.79</td>
<td>6.63</td>
<td>6.08</td>
<td>7.83</td>
<td>6.59</td>
</tr>
<tr>
<td>.03</td>
<td>5.65</td>
<td>8.26</td>
<td>6.62</td>
<td>7.16</td>
<td>6.45</td>
<td>7.92</td>
<td>6.61</td>
</tr>
<tr>
<td>.05</td>
<td>6.71</td>
<td>8.23</td>
<td>6.64</td>
<td>8.02</td>
<td>7.00</td>
<td>8.28</td>
<td>6.80</td>
</tr>
<tr>
<td>.07</td>
<td>7.77</td>
<td>8.31</td>
<td>6.67</td>
<td>8.56</td>
<td>7.17</td>
<td>8.43</td>
<td>6.80</td>
</tr>
<tr>
<td>.10</td>
<td>9.51</td>
<td>8.44</td>
<td>6.75</td>
<td>8.98</td>
<td>7.16</td>
<td>8.58</td>
<td>6.82</td>
</tr>
<tr>
<td>.14</td>
<td>11.60</td>
<td>8.48</td>
<td>6.74</td>
<td>8.95</td>
<td>6.91</td>
<td>8.55</td>
<td>6.76</td>
</tr>
<tr>
<td>.17</td>
<td>13.34</td>
<td>8.08</td>
<td>6.45</td>
<td>8.40</td>
<td>6.54</td>
<td>8.11</td>
<td>6.46</td>
</tr>
<tr>
<td>.20</td>
<td>15.24</td>
<td>8.15</td>
<td>6.54</td>
<td>8.29</td>
<td>6.57</td>
<td>8.16</td>
<td>6.54</td>
</tr>
<tr>
<td>.25</td>
<td>17.81</td>
<td>8.28</td>
<td>6.63</td>
<td>8.33</td>
<td>6.63</td>
<td>8.28</td>
<td>6.63</td>
</tr>
<tr>
<td>.30</td>
<td>20.78</td>
<td>7.96</td>
<td>6.38</td>
<td>7.99</td>
<td>6.38</td>
<td>7.96</td>
<td>6.38</td>
</tr>
<tr>
<td>.40</td>
<td>26.21</td>
<td>7.92</td>
<td>6.32</td>
<td>7.92</td>
<td>6.32</td>
<td>7.92</td>
<td>6.32</td>
</tr>
<tr>
<td>.50</td>
<td>31.37</td>
<td>7.84</td>
<td>6.26</td>
<td>7.84</td>
<td>6.26</td>
<td>7.84</td>
<td>6.26</td>
</tr>
</tbody>
</table>

Table 5.13 contains the mean squared error for estimators of $\beta_0$. Table 5.14 contains the mean squared error for estimators of $\beta_1$. The pretest estimators are for $\alpha = 0.05$ and $\alpha = 0.10$. The values of the mean squared error of $\hat{\beta}_{PW,0}$ are almost the same as the values of the mean squared error of $\hat{\beta}_{PS,0}$. The PS estimator doesn’t achieve any gain. The IV1 estimator is more efficient than the PW estimator and the PS estimator, because the IV1 estimator contains more instrumental variables than the PW estimator and the PS estimator. The IV2 estimators always have smaller mean squared error than the IV1 estimators. The mean squared errors of two pretest estimators are between the mean squared error of the OLS estimator and the mean squared errors of two IV estimators. As $\psi$ gets larger, the mean squared errors of two pretest estimators become closer to the mean squared errors of two IV estimators. The reason for this trend is that the pretest procedure rejects the null hypothesis more frequently when the correlation
Table 5.11: Monte Carlo Probability that $|t_{\hat{\beta}_0}| > t_{.025}$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{pre,PW}$</th>
<th>$\hat{\beta}_{pre,PS}$</th>
<th>$\hat{\beta}_{IV1}$</th>
<th>$\hat{\beta}_{IV2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>0.051</td>
<td>0.058</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
<td>0.071</td>
<td>0.070</td>
</tr>
<tr>
<td>.0025</td>
<td>0.075</td>
<td>0.062</td>
<td>0.062</td>
<td>0.085</td>
<td>0.084</td>
<td>0.086</td>
<td>0.084</td>
</tr>
<tr>
<td>.01</td>
<td>0.136</td>
<td>0.060</td>
<td>0.062</td>
<td>0.139</td>
<td>0.136</td>
<td>0.117</td>
<td>0.114</td>
</tr>
<tr>
<td>.02</td>
<td>0.236</td>
<td>0.064</td>
<td>0.064</td>
<td>0.214</td>
<td>0.203</td>
<td>0.152</td>
<td>0.139</td>
</tr>
<tr>
<td>.03</td>
<td>0.316</td>
<td>0.060</td>
<td>0.061</td>
<td>0.255</td>
<td>0.230</td>
<td>0.162</td>
<td>0.141</td>
</tr>
<tr>
<td>.05</td>
<td>0.486</td>
<td>0.063</td>
<td>0.064</td>
<td>0.305</td>
<td>0.249</td>
<td>0.160</td>
<td>0.124</td>
</tr>
<tr>
<td>.07</td>
<td>0.614</td>
<td>0.063</td>
<td>0.063</td>
<td>0.289</td>
<td>0.207</td>
<td>0.133</td>
<td>0.098</td>
</tr>
<tr>
<td>.10</td>
<td>0.769</td>
<td>0.062</td>
<td>0.062</td>
<td>0.221</td>
<td>0.139</td>
<td>0.095</td>
<td>0.074</td>
</tr>
<tr>
<td>.14</td>
<td>0.894</td>
<td>0.061</td>
<td>0.062</td>
<td>0.139</td>
<td>0.085</td>
<td>0.071</td>
<td>0.065</td>
</tr>
<tr>
<td>.17</td>
<td>0.941</td>
<td>0.060</td>
<td>0.061</td>
<td>0.098</td>
<td>0.069</td>
<td>0.063</td>
<td>0.062</td>
</tr>
<tr>
<td>.20</td>
<td>0.967</td>
<td>0.063</td>
<td>0.064</td>
<td>0.079</td>
<td>0.066</td>
<td>0.064</td>
<td>0.064</td>
</tr>
<tr>
<td>.25</td>
<td>0.988</td>
<td>0.060</td>
<td>0.060</td>
<td>0.064</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td>.30</td>
<td>0.996</td>
<td>0.066</td>
<td>0.066</td>
<td>0.066</td>
<td>0.066</td>
<td>0.066</td>
<td>0.066</td>
</tr>
<tr>
<td>.40</td>
<td>1.000</td>
<td>0.056</td>
<td>0.057</td>
<td>0.056</td>
<td>0.057</td>
<td>0.056</td>
<td>0.057</td>
</tr>
<tr>
<td>.50</td>
<td>1.000</td>
<td>0.064</td>
<td>0.063</td>
<td>0.064</td>
<td>0.063</td>
<td>0.064</td>
<td>0.063</td>
</tr>
</tbody>
</table>

between $p_i$ and $e_i$ increases.

As the simulation results of Table 5.15 illustrate, almost all statistics $t_{\hat{\beta}_{ols,0}}$, $t_{\hat{\beta}_{PW,0}}$, $t_{\hat{\beta}_{PS,0}}$, $t_{\hat{\beta}_{IV1,0}}$, $t_{\hat{\beta}_{IV2,0}}$ and $t_{\hat{\beta}_{pre,0}}$ exceed the tabular $t_{.025}$ for Student’s $t$ by more than 0.05. As $\psi$ increases, the probabilities $P(|t_{\hat{\beta}_{pre,0}}| > t_{.025})$ are closer to the probabilities $P(|t_{\hat{\beta}_{IV1,0}}| > t_{.025})$ and the probabilities $P(|t_{\hat{\beta}_{IV2,0}}| > t_{.025})$. Table 5.16 gives the probabilities of the $t$ statistics for $\beta$ exceeding the tabular $t_{.025}$. We can see the same tendency in Table 5.16 as in Table 5.15.

Figure 5.4 is the plot of the mean squared errors of $\hat{\beta}_{ols,0}$, $\hat{\beta}_{IV1,0}$ and $\hat{\beta}_{pre,0}$ relative to the mean squared error of $\hat{\beta}_{IV1,0}$ as a function of the correlation between $p_i$ and $e_i$ for $\alpha = 0.05$ and $\alpha = 0.10$. In Figure 5.4 the horizontal line always equal to one is the mean squared error efficiency of $\hat{\beta}_{IV1,0}$ relative to itself. The $\hat{\beta}_{ols,0}$ beats all other estimators when the correlation between $p_i$ and $e_i$ is low, but efficiency reduces when the correlation between $p_i$ and $e_i$ gets large. The curve of the IV2 estimator is always below the curve of the IV1 estimator. Two pretest estimators $\hat{\beta}_{pre,0}$ are better than $\hat{\beta}_{IV1,0}$ and
Table 5.12: Monto Carlo Probability that $|t_{\hat{\beta}}| > t_{.025}$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{pre,PW}$</th>
<th>$\hat{\beta}_{pre,PS}$</th>
<th>$\hat{\beta}_{pre,PW}$</th>
<th>$\hat{\beta}_{pre,PS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.051</td>
<td>0.074</td>
<td>0.070</td>
<td>0.061</td>
<td>0.059</td>
<td>0.081</td>
<td>0.073</td>
</tr>
<tr>
<td>0.0025</td>
<td>0.057</td>
<td>0.077</td>
<td>0.071</td>
<td>0.067</td>
<td>0.066</td>
<td>0.081</td>
<td>0.075</td>
</tr>
<tr>
<td>0.01</td>
<td>0.064</td>
<td>0.074</td>
<td>0.070</td>
<td>0.076</td>
<td>0.075</td>
<td>0.086</td>
<td>0.081</td>
</tr>
<tr>
<td>0.02</td>
<td>0.085</td>
<td>0.074</td>
<td>0.072</td>
<td>0.093</td>
<td>0.090</td>
<td>0.091</td>
<td>0.085</td>
</tr>
<tr>
<td>0.03</td>
<td>0.102</td>
<td>0.072</td>
<td>0.067</td>
<td>0.102</td>
<td>0.095</td>
<td>0.094</td>
<td>0.082</td>
</tr>
<tr>
<td>0.05</td>
<td>0.130</td>
<td>0.073</td>
<td>0.068</td>
<td>0.111</td>
<td>0.101</td>
<td>0.093</td>
<td>0.081</td>
</tr>
<tr>
<td>0.07</td>
<td>0.163</td>
<td>0.070</td>
<td>0.067</td>
<td>0.114</td>
<td>0.097</td>
<td>0.086</td>
<td>0.075</td>
</tr>
<tr>
<td>0.10</td>
<td>0.223</td>
<td>0.078</td>
<td>0.071</td>
<td>0.114</td>
<td>0.091</td>
<td>0.089</td>
<td>0.075</td>
</tr>
<tr>
<td>0.14</td>
<td>0.281</td>
<td>0.082</td>
<td>0.075</td>
<td>0.106</td>
<td>0.083</td>
<td>0.086</td>
<td>0.076</td>
</tr>
<tr>
<td>0.17</td>
<td>0.336</td>
<td>0.077</td>
<td>0.067</td>
<td>0.093</td>
<td>0.070</td>
<td>0.079</td>
<td>0.067</td>
</tr>
<tr>
<td>0.20</td>
<td>0.393</td>
<td>0.078</td>
<td>0.070</td>
<td>0.086</td>
<td>0.072</td>
<td>0.079</td>
<td>0.070</td>
</tr>
<tr>
<td>0.25</td>
<td>0.458</td>
<td>0.082</td>
<td>0.074</td>
<td>0.085</td>
<td>0.074</td>
<td>0.083</td>
<td>0.074</td>
</tr>
<tr>
<td>0.30</td>
<td>0.534</td>
<td>0.077</td>
<td>0.072</td>
<td>0.077</td>
<td>0.072</td>
<td>0.077</td>
<td>0.072</td>
</tr>
<tr>
<td>0.40</td>
<td>0.644</td>
<td>0.079</td>
<td>0.072</td>
<td>0.079</td>
<td>0.072</td>
<td>0.079</td>
<td>0.072</td>
</tr>
<tr>
<td>0.50</td>
<td>0.735</td>
<td>0.083</td>
<td>0.073</td>
<td>0.083</td>
<td>0.073</td>
<td>0.083</td>
<td>0.073</td>
</tr>
</tbody>
</table>

$\hat{\beta}_{IV2,0}$, but worse than $\hat{\beta}_{ols,0}$ when the correlation between $p_i$ and $e_i$ is low. But when the correlation is in the interval [0.1, 0.35], the two $\hat{\beta}_{pre,0}$ are worse than $\hat{\beta}_{IV1,0}$ and $\hat{\beta}_{IV2,0}$, but better than $\hat{\beta}_{ols,0}$. When the correlation keeps increasing, the two $\hat{\beta}_{pre,0}$ are worse than $\hat{\beta}_{IV2,0}$, but better than $\hat{\beta}_{IV1,0}$. The curve of the pretest estimator with $\alpha = 0.10$ is flatter than the curve of the pretest estimator with $\alpha = 0.05$.

Figure 5.5 is the plot of the mean squared errors of $\hat{\beta}_{ols,1}$, $\hat{\beta}_{IV2,1}$ and $\hat{\beta}_{pre,1}$ relative to the mean squared error of $\hat{\beta}_{IV1,1}$ for $\alpha = 0.05$ and $\alpha = 0.10$. The pretest estimators are always superior to the IV1 estimator because the correlation between $p_ix_i$ and $e_i$ is zero for all parameter sets represented in this plot. If we changed the x-axis to be the correlation between $p_ix_i$ and $e_i$, we would see the typical pretest procedure shape.
Table 5.13: Monte Carlo Mean Squared Error (×1000) for estimators of $\beta_0$ (10,000 samples)

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{IV1}$</th>
<th>$\hat{\beta}_{IV2}$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.05$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.00</td>
<td>2.29</td>
<td>5.67</td>
<td>5.68</td>
<td>5.47</td>
<td>5.11</td>
<td>2.89</td>
<td>3.26</td>
</tr>
<tr>
<td>.0025</td>
<td>3.43</td>
<td>5.85</td>
<td>5.86</td>
<td>5.65</td>
<td>5.25</td>
<td>4.12</td>
<td>4.45</td>
</tr>
<tr>
<td>.01</td>
<td>6.59</td>
<td>5.93</td>
<td>5.92</td>
<td>5.75</td>
<td>5.27</td>
<td>7.04</td>
<td>6.92</td>
</tr>
<tr>
<td>.02</td>
<td>10.96</td>
<td>5.77</td>
<td>5.77</td>
<td>5.60</td>
<td>5.20</td>
<td>9.82</td>
<td>8.97</td>
</tr>
<tr>
<td>.03</td>
<td>14.93</td>
<td>5.72</td>
<td>5.72</td>
<td>5.55</td>
<td>5.13</td>
<td>11.20</td>
<td>9.74</td>
</tr>
<tr>
<td>.05</td>
<td>23.62</td>
<td>5.48</td>
<td>5.48</td>
<td>5.32</td>
<td>4.92</td>
<td>11.46</td>
<td>9.25</td>
</tr>
<tr>
<td>.07</td>
<td>32.43</td>
<td>5.70</td>
<td>5.70</td>
<td>5.53</td>
<td>5.11</td>
<td>10.39</td>
<td>8.34</td>
</tr>
<tr>
<td>.10</td>
<td>45.05</td>
<td>5.60</td>
<td>5.61</td>
<td>5.43</td>
<td>5.05</td>
<td>7.75</td>
<td>6.48</td>
</tr>
<tr>
<td>.14</td>
<td>63.15</td>
<td>5.62</td>
<td>5.62</td>
<td>5.47</td>
<td>5.07</td>
<td>5.92</td>
<td>5.55</td>
</tr>
<tr>
<td>.17</td>
<td>75.52</td>
<td>5.58</td>
<td>5.58</td>
<td>5.46</td>
<td>5.08</td>
<td>5.47</td>
<td>5.37</td>
</tr>
<tr>
<td>.20</td>
<td>88.48</td>
<td>5.49</td>
<td>5.49</td>
<td>5.35</td>
<td>5.01</td>
<td>5.24</td>
<td>5.21</td>
</tr>
<tr>
<td>.25</td>
<td>110.17</td>
<td>5.62</td>
<td>5.63</td>
<td>5.51</td>
<td>5.17</td>
<td>5.32</td>
<td>5.34</td>
</tr>
<tr>
<td>.30</td>
<td>130.85</td>
<td>5.40</td>
<td>5.41</td>
<td>5.29</td>
<td>4.96</td>
<td>5.06</td>
<td>5.12</td>
</tr>
<tr>
<td>.40</td>
<td>174.24</td>
<td>5.30</td>
<td>5.30</td>
<td>5.23</td>
<td>4.90</td>
<td>5.01</td>
<td>5.07</td>
</tr>
<tr>
<td>.50</td>
<td>217.14</td>
<td>5.24</td>
<td>5.24</td>
<td>5.20</td>
<td>4.91</td>
<td>5.00</td>
<td>5.04</td>
</tr>
</tbody>
</table>

5.4 Concluding Remarks

The results of case one study show that when the correlation between the selection probabilities and the explanatory variables is low, then the PS estimator and the PW estimator nearly coincide. When the correlation between the selection probabilities and the explanatory variables is fairly high, the use of the $q$-weights in the PS estimator is more efficient for estimating the regression coefficients than the use of the sampling weights in the PW estimator.

The results of case two study show that if the selection probabilities are correlated with $e_i$, it’s necessary to use sampling weights. The pretest estimator based on the test for importance of weights is a compromise between the unweighted estimator and weighted estimator.

The results of case three study show under the regression model with endogenous
explanatory variables, the ordinary least squares estimator can have large bias. The instrumental variable estimator provides consistent estimators for the regression coefficients. The pretest estimator based on the test of endogeneity is a compromise between alternative instrumental variable estimators.
Table 5.15: Monte Carlo Probability that $|t_{\hat{\beta}_0}| > t_{0.025} (10,000 samples)$

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{IV1}$</th>
<th>$\hat{\beta}_{IV2}$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.05$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.049</td>
<td>0.058</td>
<td>0.058</td>
<td>0.057</td>
<td>0.053</td>
<td>0.059</td>
<td>0.065</td>
</tr>
<tr>
<td>0.025</td>
<td>0.108</td>
<td>0.061</td>
<td>0.061</td>
<td>0.063</td>
<td>0.059</td>
<td>0.117</td>
<td>0.119</td>
</tr>
<tr>
<td>0.01</td>
<td>0.282</td>
<td>0.065</td>
<td>0.065</td>
<td>0.066</td>
<td>0.061</td>
<td>0.258</td>
<td>0.237</td>
</tr>
<tr>
<td>0.02</td>
<td>0.486</td>
<td>0.061</td>
<td>0.060</td>
<td>0.061</td>
<td>0.057</td>
<td>0.373</td>
<td>0.320</td>
</tr>
<tr>
<td>0.03</td>
<td>0.649</td>
<td>0.059</td>
<td>0.059</td>
<td>0.061</td>
<td>0.059</td>
<td>0.412</td>
<td>0.330</td>
</tr>
<tr>
<td>0.05</td>
<td>0.870</td>
<td>0.055</td>
<td>0.055</td>
<td>0.056</td>
<td>0.051</td>
<td>0.342</td>
<td>0.247</td>
</tr>
<tr>
<td>0.07</td>
<td>0.950</td>
<td>0.065</td>
<td>0.065</td>
<td>0.065</td>
<td>0.062</td>
<td>0.234</td>
<td>0.167</td>
</tr>
<tr>
<td>0.10</td>
<td>0.990</td>
<td>0.059</td>
<td>0.059</td>
<td>0.059</td>
<td>0.055</td>
<td>0.116</td>
<td>0.086</td>
</tr>
<tr>
<td>0.14</td>
<td>0.999</td>
<td>0.060</td>
<td>0.059</td>
<td>0.060</td>
<td>0.055</td>
<td>0.065</td>
<td>0.062</td>
</tr>
<tr>
<td>0.17</td>
<td>1.000</td>
<td>0.062</td>
<td>0.062</td>
<td>0.064</td>
<td>0.061</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>0.20</td>
<td>1.000</td>
<td>0.058</td>
<td>0.059</td>
<td>0.060</td>
<td>0.055</td>
<td>0.059</td>
<td>0.059</td>
</tr>
<tr>
<td>0.25</td>
<td>1.000</td>
<td>0.063</td>
<td>0.064</td>
<td>0.066</td>
<td>0.060</td>
<td>0.063</td>
<td>0.064</td>
</tr>
<tr>
<td>0.30</td>
<td>1.000</td>
<td>0.059</td>
<td>0.060</td>
<td>0.064</td>
<td>0.059</td>
<td>0.062</td>
<td>0.063</td>
</tr>
<tr>
<td>0.40</td>
<td>1.000</td>
<td>0.063</td>
<td>0.063</td>
<td>0.066</td>
<td>0.060</td>
<td>0.063</td>
<td>0.065</td>
</tr>
<tr>
<td>0.50</td>
<td>1.000</td>
<td>0.060</td>
<td>0.060</td>
<td>0.064</td>
<td>0.060</td>
<td>0.063</td>
<td>0.065</td>
</tr>
</tbody>
</table>

Table 5.16: Monte Carlo Probability that $|t_{\hat{\beta}_1}| > t_{0.025} (10,000 samples)$

<table>
<thead>
<tr>
<th>$\psi$</th>
<th>$\hat{\beta}_{ols}$</th>
<th>$\hat{\beta}_{PW}$</th>
<th>$\hat{\beta}_{PS}$</th>
<th>$\hat{\beta}_{IV1}$</th>
<th>$\hat{\beta}_{IV2}$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.05$</th>
<th>$\hat{\beta}_{pre}$ $\alpha = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.049</td>
<td>0.069</td>
<td>0.069</td>
<td>0.066</td>
<td>0.051</td>
<td>0.058</td>
<td>0.063</td>
</tr>
<tr>
<td>0.025</td>
<td>0.052</td>
<td>0.067</td>
<td>0.069</td>
<td>0.066</td>
<td>0.057</td>
<td>0.060</td>
<td>0.067</td>
</tr>
<tr>
<td>0.01</td>
<td>0.054</td>
<td>0.073</td>
<td>0.073</td>
<td>0.070</td>
<td>0.055</td>
<td>0.065</td>
<td>0.072</td>
</tr>
<tr>
<td>0.02</td>
<td>0.057</td>
<td>0.073</td>
<td>0.072</td>
<td>0.068</td>
<td>0.053</td>
<td>0.067</td>
<td>0.074</td>
</tr>
<tr>
<td>0.03</td>
<td>0.062</td>
<td>0.074</td>
<td>0.074</td>
<td>0.071</td>
<td>0.057</td>
<td>0.073</td>
<td>0.079</td>
</tr>
<tr>
<td>0.05</td>
<td>0.072</td>
<td>0.070</td>
<td>0.071</td>
<td>0.065</td>
<td>0.056</td>
<td>0.072</td>
<td>0.080</td>
</tr>
<tr>
<td>0.07</td>
<td>0.077</td>
<td>0.070</td>
<td>0.072</td>
<td>0.068</td>
<td>0.057</td>
<td>0.078</td>
<td>0.085</td>
</tr>
<tr>
<td>0.10</td>
<td>0.083</td>
<td>0.073</td>
<td>0.072</td>
<td>0.069</td>
<td>0.054</td>
<td>0.074</td>
<td>0.081</td>
</tr>
<tr>
<td>0.14</td>
<td>0.100</td>
<td>0.072</td>
<td>0.072</td>
<td>0.069</td>
<td>0.050</td>
<td>0.072</td>
<td>0.078</td>
</tr>
<tr>
<td>0.17</td>
<td>0.112</td>
<td>0.077</td>
<td>0.076</td>
<td>0.069</td>
<td>0.055</td>
<td>0.072</td>
<td>0.079</td>
</tr>
<tr>
<td>0.20</td>
<td>0.119</td>
<td>0.071</td>
<td>0.070</td>
<td>0.067</td>
<td>0.052</td>
<td>0.068</td>
<td>0.074</td>
</tr>
<tr>
<td>0.25</td>
<td>0.144</td>
<td>0.072</td>
<td>0.072</td>
<td>0.069</td>
<td>0.055</td>
<td>0.071</td>
<td>0.077</td>
</tr>
<tr>
<td>0.30</td>
<td>0.154</td>
<td>0.076</td>
<td>0.077</td>
<td>0.072</td>
<td>0.053</td>
<td>0.069</td>
<td>0.076</td>
</tr>
<tr>
<td>0.40</td>
<td>0.197</td>
<td>0.074</td>
<td>0.074</td>
<td>0.073</td>
<td>0.059</td>
<td>0.070</td>
<td>0.076</td>
</tr>
<tr>
<td>0.50</td>
<td>0.233</td>
<td>0.074</td>
<td>0.075</td>
<td>0.072</td>
<td>0.054</td>
<td>0.066</td>
<td>0.070</td>
</tr>
</tbody>
</table>
Figure 5.2: Plot of MSEs relative to that of $\hat{\beta}_{PS,0}$
Figure 5.3: Plot of MSEs relative to that of $\hat{\beta}_{PS,1}$
Figure 5.4: Plot of MSEs relative to that of $\hat{\beta}_{IV_{0,0}}$
Figure 5.5: Plot of MSEs relative to that of $\hat{\beta}_{IV_{0.1}}$
5.5 A Sample Example

To illustrate computation of different estimation procedures, a sample was constructed by the procedure described in Section 5.1 with \(a = 0.25\) and \(b = 1.75\) in (5.1). In the population \(\text{corr}(p_i, x_i) = 0.559\), \(\text{corr}(p_i, e_i) = 0.104\), and \(\text{corr}(p_i x_i, e_i) = 0.011\). The procedure gave a sample with size 231. The data set is shown in Table 5.17. The simple sample mean vector is \((\bar{y}_n, \bar{x}_n) = (1.093, 0.066)\) and the weighted mean vector is \((\bar{y}_\pi, \bar{x}_\pi) = (0.641, 0.044)\), where \(\bar{y}_\pi = \frac{1}{\sum_{i=1}^{n} \pi_i} \sum_{i=1}^{n} \pi_i^{-1} y_i\) and \(\bar{x}_\pi = \frac{1}{\sum_{i=1}^{n} \pi_i^{-1}} \sum_{i=1}^{n} \pi_i^{-1} x_i\). The simple sample means and the weighted sample means are different because of the positive correlation between \(p_i\) and \(e_i\) and the positive correlation between \(p_i\) and \(x_i\).

Fitting the regression by ordinary least squares (OLS), we obtain

\[
\hat{y}_{i,\text{ols}} = 1.031 + 0.940 x_i \quad (5.19)
\]

\[
(0.047) (0.065)
\]

\(s^2 = 0.499\),

where \(\bar{x}_n = 0.066\) is the simple sample mean, and the numbers in parenthesis are the standard errors from the ordinary least squares calculations. The estimated standard errors are incorrect, because \(\pi_i\) and \(e_i\) are correlated, and the distribution of the sample observations is not the same as the distribution of the population elements.

The probability weighted regression of (1.8) is

\[
\hat{y}_{i,\text{PW}} = 0.601 + 0.934 x_i \quad (5.20)
\]

\[
(0.060) (0.079)
\]

where the numbers in parenthesis are the standard errors calculated using the covariance matrix (1.12). The estimated covariance matrix is

\[
\hat{V}(\hat{\beta}_{PW}) = \begin{pmatrix}
3.355 & -0.624 \\
-0.624 & 6.267
\end{pmatrix} 10^{-3}.
\]
The estimated intercepts in the equations (5.19) and (5.20) are greatly different, because the bias for the intercept mainly comes from the positive correlation between $p_i$ and $e_i$. The estimated slopes in the equations (5.19) and (5.20) are almost the same due to the weak correlation between $p_i$ and $x_i$ and the weak correlation between $p_i x_i$ and $e_i$. The standard errors of the weighted estimator are larger than those of the OLS.

To test the hypothesis that the two estimators are estimating the same quantity we compute the OLS regression of $y_i$ on $(1, x_i, w_i, w_i x_i)$, where $w_i = p_i^{-1}$ is sampling weight. The estimated equation is

$$\hat{y}_i = 1.355 + 1.103 x_i - 0.071 w_i - 0.051 w_i x_i$$

$$s^2 = 0.353,$$

where the numbers in parenthesis are the standard errors calculated from the OLS calculations. The $F$ test for the hypothesis of common expectation for $\hat{\beta}_{PW}$ and $\hat{\beta}_{ols}$ is

$$F_{227}^2 = 0.5(114.306 - 80.917)(0.356)^{-1} = 46.83,$$

The ten percent point for $F_{227}^2$ is 2.30 and we conclude that the ordinary least squares estimator is biased for $\beta_0$ and $\beta_1$.

In computing the PS estimators, estimated probabilities $\hat{p}_i$'s are constructed, where $\hat{p}_i$ is the predicted value from the OLS regression of $p_i$ on $(1, r(x_i))$, and the $q$-weight in the PS estimator defined in (2.32) is $q_i = w_i \hat{p}_i$. The estimated equation for the PS estimator is

$$\hat{y}_{i, PS} = 0.600 + 0.937 x_i,$$

where the standard errors were calculated using equation (5.4). The estimated covariance
The matrix is
\[ \hat{V}(\hat{\beta}_{PS}) = \begin{pmatrix} 3.481 & -0.711 \\ -0.711 & 6.020 \end{pmatrix} \times 10^{-3}. \]

The equations (5.20) and (5.22) are not greatly different.

We construct two IV estimators. The first IV estimator is based on four instrumental variables, \( w_i, w_i x_i, q_i, \) and \( q_i x_i \). The second IV estimator is based on five instrumental variables, \( w_i, w_i x_i, q_i, q_i x_i, \) and \( x_i \). The first IV estimator is defined in (5.5) with \( z_{2,i} = (w_i, w_i x_i, q_i, q_i x_i) \), \( Z_2 = (z_{2,1}, z_{2,2}, \ldots, z_{2,n})' \) is an \( n \times 4 \) matrix and \( q_i = w_i \hat{w}_i^{-1} \). The estimated equation is
\[ \hat{y}_{i,IV1} = 0.596 + 0.944 x_i, \quad (5.23) \]
\[ (0.058) (0.076) \]

where the standard errors were calculated using equation (5.6). The estimated covariance matrix is
\[ \hat{V}(\hat{\beta}_{IV1}) = \begin{pmatrix} 3.337 & -0.583 \\ -0.583 & 5.778 \end{pmatrix} \times 10^{-3}. \]

The second IV estimator is defined in (5.7) with \( z_i = (w_i, w_i x_i, q_i, q_i x_i, z_{3,i}) \), \( z_{3,i} = x_i \), \( z_3 = (z_{3,1}, z_{3,2}, \ldots, z_{3,n})' \), and \( Z = (Z_2, z_3) \) is an \( n \times 5 \) matrix. The estimated equation is
\[ \hat{y}_{i,IV2} = 0.579 + 0.998 x_i, \quad (5.24) \]
\[ (0.059) (0.072) \]

where the standard errors were calculated using equation (5.8). The estimated covariance matrix is
\[ \hat{V}(\hat{\beta}_{IV2}) = \begin{pmatrix} 3.433 & -0.559 \\ -0.559 & 5.245 \end{pmatrix} \times 10^{-3}. \]

The equations (5.23) and (5.24) are slightly different. The standard error of the slope of the IV2 estimator are slightly less than that of IV1 estimator.
To test the hypothesis that $x_i$ can be an instrumental variable we construct a test for endogeneity. We compute the OLS regression of $y_i$ on $(\bar{x}_0, \bar{x}_i, z_{3,i} - \hat{z}_{3,i})$, where $\bar{x}_0$ is the predicted value from the regression of 1 on $(w_i, w_i x_i, q_i, q_i x_i, z_{3,i})$, $\bar{x}_i$ is the predicted value from the regression of $x_i$ on $(w_i, w_i x_i, q_i, q_i x_i, z_{3,i})$, and $\hat{z}_{3,i}$ is the predicted value from the regression of $z_{3,i}$ on $(w_i, w_i x_i, q_i, q_i x_i)$, where $z_{3,i} = x_i$. The estimated equation is

$$\hat{y}_i = 0.596 \bar{x}_0 + 0.944 \bar{x}_i + 0.106(z_{3,i} - \hat{z}_{3,i})$$

(5.25)

and the $t$ test for the hypothesis $H_0: \delta = 0$ is

$$t = 0.106/0.124 = 0.86.$$  

The value of the $t$-statistic is less than 1.65, thus we can use $x_i$ as an instrumental variable.
### Table 5.17: Example Sample Data

<table>
<thead>
<tr>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.3243</td>
<td>0.3796</td>
<td>1.5173</td>
<td>2</td>
<td>1.3254</td>
<td>0.5253</td>
<td>1.2336</td>
</tr>
<tr>
<td>3</td>
<td>1.3247</td>
<td>-0.4123</td>
<td>1.5828</td>
<td>4</td>
<td>6.8373</td>
<td>-0.7427</td>
<td>-0.4050</td>
</tr>
<tr>
<td>5</td>
<td>1.6304</td>
<td>-0.2071</td>
<td>0.4816</td>
<td>6</td>
<td>2.0002</td>
<td>0.3260</td>
<td>0.8365</td>
</tr>
<tr>
<td>7</td>
<td>2.8343</td>
<td>-0.1712</td>
<td>-0.7108</td>
<td>8</td>
<td>1.1348</td>
<td>0.0059</td>
<td>0.7456</td>
</tr>
<tr>
<td>9</td>
<td>1.1348</td>
<td>-0.2067</td>
<td>1.5525</td>
<td>10</td>
<td>1.1683</td>
<td>-0.2956</td>
<td>0.9302</td>
</tr>
<tr>
<td>11</td>
<td>1.2318</td>
<td>-0.3230</td>
<td>1.4963</td>
<td>12</td>
<td>20.0000</td>
<td>-0.2076</td>
<td>-0.3101</td>
</tr>
<tr>
<td>13</td>
<td>3.3372</td>
<td>0.4393</td>
<td>0.7860</td>
<td>14</td>
<td>1.2992</td>
<td>1.3837</td>
<td>2.1939</td>
</tr>
<tr>
<td>15</td>
<td>9.1789</td>
<td>0.6964</td>
<td>0.7711</td>
<td>16</td>
<td>1.3218</td>
<td>-1.1388</td>
<td>0.2089</td>
</tr>
<tr>
<td>17</td>
<td>1.1348</td>
<td>0.1678</td>
<td>1.8656</td>
<td>18</td>
<td>2.4871</td>
<td>-0.5164</td>
<td>0.2447</td>
</tr>
<tr>
<td>19</td>
<td>1.1348</td>
<td>0.3477</td>
<td>0.8528</td>
<td>20</td>
<td>1.4548</td>
<td>-0.7915</td>
<td>0.7831</td>
</tr>
<tr>
<td>21</td>
<td>1.1348</td>
<td>0.6984</td>
<td>1.7022</td>
<td>22</td>
<td>8.0788</td>
<td>0.8213</td>
<td>0.1439</td>
</tr>
<tr>
<td>23</td>
<td>1.1016</td>
<td>1.2176</td>
<td>2.3932</td>
<td>24</td>
<td>1.0715</td>
<td>0.6381</td>
<td>2.0563</td>
</tr>
<tr>
<td>25</td>
<td>1.1198</td>
<td>0.2991</td>
<td>2.0427</td>
<td>26</td>
<td>4.6756</td>
<td>1.2558</td>
<td>1.3769</td>
</tr>
<tr>
<td>27</td>
<td>3.0441</td>
<td>-0.7859</td>
<td>-0.0480</td>
<td>28</td>
<td>6.8647</td>
<td>0.0387</td>
<td>1.3561</td>
</tr>
<tr>
<td>29</td>
<td>20.0000</td>
<td>-0.2847</td>
<td>1.3238</td>
<td>30</td>
<td>3.3726</td>
<td>-0.3653</td>
<td>-0.3874</td>
</tr>
<tr>
<td>31</td>
<td>1.8918</td>
<td>-1.1501</td>
<td>0.0963</td>
<td>32</td>
<td>1.5079</td>
<td>-1.3453</td>
<td>-0.5618</td>
</tr>
<tr>
<td>33</td>
<td>1.6683</td>
<td>-0.3530</td>
<td>0.6640</td>
<td>34</td>
<td>1.1734</td>
<td>-0.3938</td>
<td>0.0940</td>
</tr>
<tr>
<td>35</td>
<td>1.2276</td>
<td>0.6940</td>
<td>2.0127</td>
<td>36</td>
<td>20.0000</td>
<td>-0.3384</td>
<td>0.7991</td>
</tr>
<tr>
<td>37</td>
<td>1.0990</td>
<td>0.4413</td>
<td>2.2512</td>
<td>38</td>
<td>2.0084</td>
<td>-0.7058</td>
<td>-0.7538</td>
</tr>
<tr>
<td>39</td>
<td>9.1908</td>
<td>0.6952</td>
<td>1.3861</td>
<td>40</td>
<td>3.6692</td>
<td>-0.1506</td>
<td>1.2758</td>
</tr>
<tr>
<td>41</td>
<td>1.0615</td>
<td>0.7123</td>
<td>2.2675</td>
<td>42</td>
<td>1.5777</td>
<td>0.3372</td>
<td>1.7191</td>
</tr>
<tr>
<td>43</td>
<td>2.4670</td>
<td>0.5040</td>
<td>1.3623</td>
<td>44</td>
<td>3.5722</td>
<td>-0.6208</td>
<td>0.4759</td>
</tr>
<tr>
<td>45</td>
<td>1.1348</td>
<td>0.0752</td>
<td>2.3182</td>
<td>46</td>
<td>2.0323</td>
<td>-0.2105</td>
<td>0.1142</td>
</tr>
<tr>
<td>47</td>
<td>1.9282</td>
<td>-0.1489</td>
<td>1.3531</td>
<td>48</td>
<td>20.0000</td>
<td>-0.7743</td>
<td>-0.6439</td>
</tr>
<tr>
<td>49</td>
<td>3.1777</td>
<td>0.5319</td>
<td>0.7589</td>
<td>50</td>
<td>13.2362</td>
<td>0.4152</td>
<td>0.6590</td>
</tr>
<tr>
<td>51</td>
<td>3.9033</td>
<td>-0.1254</td>
<td>0.9328</td>
<td>52</td>
<td>3.1041</td>
<td>-1.0253</td>
<td>-0.9793</td>
</tr>
<tr>
<td>53</td>
<td>1.1933</td>
<td>0.2163</td>
<td>1.1266</td>
<td>54</td>
<td>1.1325</td>
<td>1.5926</td>
<td>2.3028</td>
</tr>
<tr>
<td>55</td>
<td>1.1348</td>
<td>-0.5167</td>
<td>1.3188</td>
<td>56</td>
<td>5.1520</td>
<td>-0.6797</td>
<td>0.0215</td>
</tr>
<tr>
<td>57</td>
<td>1.4823</td>
<td>0.6451</td>
<td>1.6081</td>
<td>58</td>
<td>3.1520</td>
<td>0.4878</td>
<td>1.6643</td>
</tr>
<tr>
<td>59</td>
<td>2.2289</td>
<td>0.0142</td>
<td>1.4081</td>
<td>60</td>
<td>1.1370</td>
<td>0.6890</td>
<td>1.9698</td>
</tr>
<tr>
<td>61</td>
<td>1.3601</td>
<td>1.4884</td>
<td>3.4342</td>
<td>62</td>
<td>2.0890</td>
<td>-0.0014</td>
<td>1.1081</td>
</tr>
<tr>
<td>63</td>
<td>1.1784</td>
<td>0.3387</td>
<td>1.1840</td>
<td>64</td>
<td>1.1348</td>
<td>-0.1549</td>
<td>1.9337</td>
</tr>
<tr>
<td>65</td>
<td>10.8598</td>
<td>0.5544</td>
<td>-0.0081</td>
<td>66</td>
<td>2.5231</td>
<td>0.5316</td>
<td>0.9996</td>
</tr>
<tr>
<td>67</td>
<td>2.5675</td>
<td>-1.3826</td>
<td>-1.0916</td>
<td>68</td>
<td>1.3684</td>
<td>-0.3307</td>
<td>0.8180</td>
</tr>
<tr>
<td>69</td>
<td>1.8640</td>
<td>-0.2239</td>
<td>0.8002</td>
<td>70</td>
<td>1.1348</td>
<td>-1.4254</td>
<td>0.3195</td>
</tr>
<tr>
<td>71</td>
<td>2.4352</td>
<td>0.0051</td>
<td>1.1415</td>
<td>72</td>
<td>5.9259</td>
<td>1.6122</td>
<td>2.1379</td>
</tr>
<tr>
<td>73</td>
<td>1.3855</td>
<td>1.4637</td>
<td>2.5802</td>
<td>74</td>
<td>1.7850</td>
<td>0.5297</td>
<td>2.2768</td>
</tr>
<tr>
<td>75</td>
<td>2.5343</td>
<td>0.5554</td>
<td>0.8678</td>
<td>76</td>
<td>20.0000</td>
<td>0.1369</td>
<td>0.2029</td>
</tr>
<tr>
<td>77</td>
<td>1.1348</td>
<td>-0.9364</td>
<td>0.9828</td>
<td>78</td>
<td>1.1348</td>
<td>-1.1214</td>
<td>0.6945</td>
</tr>
<tr>
<td>79</td>
<td>13.1426</td>
<td>0.4197</td>
<td>1.7019</td>
<td>80</td>
<td>1.6254</td>
<td>1.0795</td>
<td>2.3297</td>
</tr>
</tbody>
</table>
Table 5.17: (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>12.3167</td>
<td>0.4627</td>
<td>0.0350</td>
<td>82</td>
<td>1.1348</td>
<td>-0.4517</td>
<td>1.2847</td>
</tr>
<tr>
<td>83</td>
<td>20.0000</td>
<td>-0.1149</td>
<td>0.0176</td>
<td>84</td>
<td>1.7602</td>
<td>0.7570</td>
<td>1.6462</td>
</tr>
<tr>
<td>85</td>
<td>3.2333</td>
<td>0.0691</td>
<td>0.8759</td>
<td>86</td>
<td>2.4142</td>
<td>-0.1792</td>
<td>0.9767</td>
</tr>
<tr>
<td>87</td>
<td>20.0000</td>
<td>-0.1927</td>
<td>-0.0286</td>
<td>88</td>
<td>2.6920</td>
<td>-1.1983</td>
<td>-0.5465</td>
</tr>
<tr>
<td>89</td>
<td>1.4068</td>
<td>-0.8606</td>
<td>0.0152</td>
<td>90</td>
<td>1.1348</td>
<td>1.1348</td>
<td>-1.2220</td>
</tr>
<tr>
<td>91</td>
<td>1.1994</td>
<td>-0.6360</td>
<td>0.3267</td>
<td>92</td>
<td>11.2470</td>
<td>0.5277</td>
<td>1.3052</td>
</tr>
<tr>
<td>93</td>
<td>1.1416</td>
<td>-0.6752</td>
<td>0.1496</td>
<td>94</td>
<td>2.7753</td>
<td>-0.1900</td>
<td>0.3641</td>
</tr>
<tr>
<td>95</td>
<td>1.3173</td>
<td>-0.5161</td>
<td>1.0292</td>
<td>96</td>
<td>1.7613</td>
<td>0.1348</td>
<td>1.6144</td>
</tr>
<tr>
<td>97</td>
<td>1.2836</td>
<td>0.2527</td>
<td>2.3536</td>
<td>98</td>
<td>1.1348</td>
<td>0.0131</td>
<td>1.3052</td>
</tr>
<tr>
<td>99</td>
<td>5.4387</td>
<td>1.4361</td>
<td>2.6644</td>
<td>100</td>
<td>10.1952</td>
<td>0.6049</td>
<td>1.8169</td>
</tr>
<tr>
<td>101</td>
<td>1.9880</td>
<td>0.6126</td>
<td>1.8155</td>
<td>102</td>
<td>13.4568</td>
<td>0.4047</td>
<td>1.5024</td>
</tr>
<tr>
<td>103</td>
<td>1.4132</td>
<td>-0.1838</td>
<td>0.5171</td>
<td>104</td>
<td>13.4568</td>
<td>0.4047</td>
<td>1.5024</td>
</tr>
<tr>
<td>105</td>
<td>2.9858</td>
<td>-0.4677</td>
<td>0.7949</td>
<td>106</td>
<td>1.7974</td>
<td>-0.4790</td>
<td>0.7099</td>
</tr>
<tr>
<td>107</td>
<td>1.1348</td>
<td>-0.1269</td>
<td>1.4900</td>
<td>108</td>
<td>8.2473</td>
<td>0.8000</td>
<td>1.0166</td>
</tr>
<tr>
<td>109</td>
<td>1.2313</td>
<td>0.8772</td>
<td>3.0616</td>
<td>110</td>
<td>14.4102</td>
<td>0.3633</td>
<td>1.1739</td>
</tr>
<tr>
<td>111</td>
<td>2.0939</td>
<td>-0.6755</td>
<td>0.3316</td>
<td>112</td>
<td>6.2278</td>
<td>1.1311</td>
<td>0.9060</td>
</tr>
<tr>
<td>113</td>
<td>16.9816</td>
<td>0.2748</td>
<td>0.2228</td>
<td>114</td>
<td>4.9526</td>
<td>-0.5780</td>
<td>0.4454</td>
</tr>
<tr>
<td>115</td>
<td>1.1001</td>
<td>0.6086</td>
<td>2.8080</td>
<td>116</td>
<td>20.0000</td>
<td>-0.2151</td>
<td>-0.1853</td>
</tr>
<tr>
<td>117</td>
<td>1.9767</td>
<td>-0.7337</td>
<td>-0.7949</td>
<td>118</td>
<td>1.0669</td>
<td>0.6720</td>
<td>2.7397</td>
</tr>
<tr>
<td>119</td>
<td>1.1348</td>
<td>0.1788</td>
<td>3.0239</td>
<td>120</td>
<td>1.6819</td>
<td>-0.1584</td>
<td>1.1729</td>
</tr>
<tr>
<td>121</td>
<td>1.0526</td>
<td>0.7795</td>
<td>2.6794</td>
<td>122</td>
<td>3.1484</td>
<td>-0.3755</td>
<td>0.5880</td>
</tr>
<tr>
<td>123</td>
<td>10.5249</td>
<td>0.4037</td>
<td>0.6643</td>
<td>124</td>
<td>4.9627</td>
<td>-0.7632</td>
<td>0.0473</td>
</tr>
<tr>
<td>125</td>
<td>1.2149</td>
<td>1.4775</td>
<td>3.2647</td>
<td>126</td>
<td>1.1348</td>
<td>-0.5049</td>
<td>1.5774</td>
</tr>
<tr>
<td>127</td>
<td>3.8399</td>
<td>-0.0173</td>
<td>0.7316</td>
<td>128</td>
<td>2.4458</td>
<td>-0.6012</td>
<td>0.6352</td>
</tr>
<tr>
<td>129</td>
<td>3.8270</td>
<td>0.5863</td>
<td>2.4071</td>
<td>130</td>
<td>2.0810</td>
<td>-1.6246</td>
<td>-1.1285</td>
</tr>
<tr>
<td>131</td>
<td>2.5865</td>
<td>0.0790</td>
<td>2.0437</td>
<td>132</td>
<td>1.7611</td>
<td>-1.6439</td>
<td>-1.0759</td>
</tr>
<tr>
<td>133</td>
<td>1.0510</td>
<td>0.7912</td>
<td>1.8622</td>
<td>134</td>
<td>1.1348</td>
<td>-0.3087</td>
<td>-0.1787</td>
</tr>
<tr>
<td>135</td>
<td>7.3779</td>
<td>0.9203</td>
<td>1.3137</td>
<td>136</td>
<td>1.1875</td>
<td>0.7722</td>
<td>1.6581</td>
</tr>
<tr>
<td>137</td>
<td>1.4537</td>
<td>1.3458</td>
<td>2.8163</td>
<td>138</td>
<td>2.7843</td>
<td>-0.3889</td>
<td>1.1870</td>
</tr>
<tr>
<td>139</td>
<td>1.1348</td>
<td>-0.2986</td>
<td>1.2308</td>
<td>140</td>
<td>7.7018</td>
<td>0.8723</td>
<td>0.3485</td>
</tr>
<tr>
<td>141</td>
<td>1.8896</td>
<td>0.0457</td>
<td>2.1768</td>
<td>142</td>
<td>12.1508</td>
<td>0.4720</td>
<td>0.9349</td>
</tr>
<tr>
<td>143</td>
<td>2.5480</td>
<td>-0.9359</td>
<td>0.7201</td>
<td>144</td>
<td>2.0672</td>
<td>-0.3255</td>
<td>0.8066</td>
</tr>
<tr>
<td>145</td>
<td>1.2842</td>
<td>0.4964</td>
<td>1.3587</td>
<td>146</td>
<td>1.2590</td>
<td>1.5292</td>
<td>3.0647</td>
</tr>
<tr>
<td>147</td>
<td>1.3080</td>
<td>0.0869</td>
<td>1.2371</td>
<td>148</td>
<td>2.9665</td>
<td>1.6899</td>
<td>2.2975</td>
</tr>
<tr>
<td>149</td>
<td>1.1348</td>
<td>-0.5119</td>
<td>0.8265</td>
<td>150</td>
<td>6.6982</td>
<td>-0.6618</td>
<td>0.8216</td>
</tr>
<tr>
<td>151</td>
<td>20.0000</td>
<td>-0.9830</td>
<td>-0.5553</td>
<td>152</td>
<td>2.6859</td>
<td>1.1076</td>
<td>1.8491</td>
</tr>
<tr>
<td>153</td>
<td>2.3594</td>
<td>0.6951</td>
<td>1.8799</td>
<td>154</td>
<td>2.2908</td>
<td>0.5357</td>
<td>0.9713</td>
</tr>
<tr>
<td>155</td>
<td>2.6872</td>
<td>-0.0998</td>
<td>1.1296</td>
<td>156</td>
<td>1.7763</td>
<td>0.9114</td>
<td>1.6654</td>
</tr>
<tr>
<td>157</td>
<td>2.6724</td>
<td>-0.1806</td>
<td>1.7636</td>
<td>158</td>
<td>11.2199</td>
<td>0.5295</td>
<td>1.4198</td>
</tr>
<tr>
<td>159</td>
<td>1.2405</td>
<td>1.3496</td>
<td>2.8111</td>
<td>160</td>
<td>7.1346</td>
<td>0.9593</td>
<td>1.7132</td>
</tr>
</tbody>
</table>
Table 5.17: (Continued)

<table>
<thead>
<tr>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
<th>ID</th>
<th>Weight</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>161</td>
<td>1.4733</td>
<td>-1.1428</td>
<td>0.9954</td>
<td>162</td>
<td>1.1923</td>
<td>1.1713</td>
<td>2.9777</td>
</tr>
<tr>
<td>163</td>
<td>1.2153</td>
<td>-1.2827</td>
<td>-0.6568</td>
<td>164</td>
<td>20.0000</td>
<td>-0.0462</td>
<td>0.2718</td>
</tr>
<tr>
<td>165</td>
<td>1.3163</td>
<td>-0.3700</td>
<td>0.5258</td>
<td>166</td>
<td>1.1348</td>
<td>-0.0906</td>
<td>2.1622</td>
</tr>
<tr>
<td>167</td>
<td>1.1617</td>
<td>-0.6996</td>
<td>0.8523</td>
<td>168</td>
<td>1.0377</td>
<td>0.8940</td>
<td>2.7644</td>
</tr>
<tr>
<td>169</td>
<td>1.5687</td>
<td>0.6690</td>
<td>2.1242</td>
<td>170</td>
<td>2.3617</td>
<td>0.1097</td>
<td>1.6019</td>
</tr>
<tr>
<td>171</td>
<td>20.0000</td>
<td>0.1206</td>
<td>-0.0690</td>
<td>172</td>
<td>1.1348</td>
<td>-0.2115</td>
<td>1.8272</td>
</tr>
<tr>
<td>173</td>
<td>1.0511</td>
<td>0.7903</td>
<td>2.1776</td>
<td>174</td>
<td>1.1348</td>
<td>-0.2010</td>
<td>1.6536</td>
</tr>
<tr>
<td>175</td>
<td>2.6244</td>
<td>0.2983</td>
<td>0.9540</td>
<td>176</td>
<td>2.8823</td>
<td>0.1984</td>
<td>1.5241</td>
</tr>
<tr>
<td>177</td>
<td>1.1348</td>
<td>-0.1307</td>
<td>2.0052</td>
<td>178</td>
<td>1.6063</td>
<td>0.5531</td>
<td>2.5295</td>
</tr>
<tr>
<td>179</td>
<td>9.5734</td>
<td>0.6586</td>
<td>0.2942</td>
<td>180</td>
<td>1.2846</td>
<td>-0.5444</td>
<td>0.8769</td>
</tr>
<tr>
<td>181</td>
<td>11.3700</td>
<td>0.5196</td>
<td>0.7540</td>
<td>182</td>
<td>1.2241</td>
<td>0.3861</td>
<td>1.0422</td>
</tr>
<tr>
<td>183</td>
<td>20.0000</td>
<td>0.0381</td>
<td>0.9929</td>
<td>184</td>
<td>2.3087</td>
<td>-0.3556</td>
<td>1.0834</td>
</tr>
<tr>
<td>185</td>
<td>7.3262</td>
<td>0.4839</td>
<td>1.0271</td>
<td>186</td>
<td>14.4401</td>
<td>-0.3698</td>
<td>0.1669</td>
</tr>
<tr>
<td>187</td>
<td>1.1348</td>
<td>0.1963</td>
<td>2.2353</td>
<td>188</td>
<td>1.4918</td>
<td>-0.5283</td>
<td>0.1165</td>
</tr>
<tr>
<td>189</td>
<td>20.0000</td>
<td>-0.3733</td>
<td>-0.4285</td>
<td>190</td>
<td>3.3128</td>
<td>-0.3193</td>
<td>0.8005</td>
</tr>
<tr>
<td>191</td>
<td>1.4164</td>
<td>-0.4736</td>
<td>1.2386</td>
<td>192</td>
<td>4.3998</td>
<td>-0.3694</td>
<td>0.3821</td>
</tr>
<tr>
<td>193</td>
<td>1.7197</td>
<td>-0.1619</td>
<td>1.1176</td>
<td>194</td>
<td>20.0000</td>
<td>-0.2857</td>
<td>0.1371</td>
</tr>
<tr>
<td>195</td>
<td>1.9617</td>
<td>0.4204</td>
<td>1.4782</td>
<td>196</td>
<td>1.8326</td>
<td>-0.5824</td>
<td>-0.1548</td>
</tr>
<tr>
<td>197</td>
<td>1.1348</td>
<td>-0.2133</td>
<td>1.9172</td>
<td>198</td>
<td>1.0777</td>
<td>0.5931</td>
<td>3.2816</td>
</tr>
<tr>
<td>199</td>
<td>1.1827</td>
<td>-0.0785</td>
<td>2.1651</td>
<td>200</td>
<td>2.7618</td>
<td>0.4931</td>
<td>0.5766</td>
</tr>
<tr>
<td>201</td>
<td>1.9523</td>
<td>-0.2792</td>
<td>0.6771</td>
<td>202</td>
<td>6.6633</td>
<td>0.4545</td>
<td>1.3771</td>
</tr>
<tr>
<td>203</td>
<td>2.4283</td>
<td>0.9659</td>
<td>2.9946</td>
<td>204</td>
<td>1.1724</td>
<td>-1.2409</td>
<td>0.4918</td>
</tr>
<tr>
<td>205</td>
<td>6.5910</td>
<td>-0.6008</td>
<td>0.5843</td>
<td>206</td>
<td>15.0876</td>
<td>-0.1596</td>
<td>0.2636</td>
</tr>
<tr>
<td>207</td>
<td>1.1348</td>
<td>0.0990</td>
<td>3.0626</td>
<td>208</td>
<td>1.1348</td>
<td>-0.7572</td>
<td>1.5950</td>
</tr>
<tr>
<td>209</td>
<td>2.2203</td>
<td>0.4853</td>
<td>1.3171</td>
<td>210</td>
<td>20.0000</td>
<td>-0.7882</td>
<td>-0.5929</td>
</tr>
<tr>
<td>211</td>
<td>5.9259</td>
<td>1.3510</td>
<td>1.0218</td>
<td>212</td>
<td>5.9259</td>
<td>1.8144</td>
<td>1.7860</td>
</tr>
<tr>
<td>213</td>
<td>1.3105</td>
<td>1.4407</td>
<td>3.1995</td>
<td>214</td>
<td>1.3469</td>
<td>-0.8609</td>
<td>0.1961</td>
</tr>
<tr>
<td>215</td>
<td>20.0000</td>
<td>-0.9028</td>
<td>-0.7839</td>
<td>216</td>
<td>20.0000</td>
<td>-0.3081</td>
<td>0.7325</td>
</tr>
<tr>
<td>217</td>
<td>1.1283</td>
<td>0.2424</td>
<td>1.4590</td>
<td>218</td>
<td>1.6327</td>
<td>-0.7474</td>
<td>0.8347</td>
</tr>
<tr>
<td>219</td>
<td>2.0771</td>
<td>-0.7743</td>
<td>0.4279</td>
<td>220</td>
<td>1.2142</td>
<td>-1.2278</td>
<td>-0.6439</td>
</tr>
<tr>
<td>221</td>
<td>2.9269</td>
<td>0.5849</td>
<td>0.9022</td>
<td>222</td>
<td>1.3033</td>
<td>0.2372</td>
<td>1.4901</td>
</tr>
<tr>
<td>223</td>
<td>6.6590</td>
<td>0.6384</td>
<td>1.4848</td>
<td>224</td>
<td>6.2263</td>
<td>0.6089</td>
<td>-0.1887</td>
</tr>
<tr>
<td>225</td>
<td>20.0000</td>
<td>-0.1987</td>
<td>0.1873</td>
<td>226</td>
<td>1.9652</td>
<td>0.0211</td>
<td>1.1264</td>
</tr>
<tr>
<td>227</td>
<td>2.2862</td>
<td>0.1604</td>
<td>1.7925</td>
<td>228</td>
<td>1.6288</td>
<td>-0.4977</td>
<td>0.3909</td>
</tr>
<tr>
<td>229</td>
<td>20.0000</td>
<td>-0.3415</td>
<td>-0.5361</td>
<td>230</td>
<td>1.1149</td>
<td>1.2605</td>
<td>1.9173</td>
</tr>
<tr>
<td>231</td>
<td>5.5039</td>
<td>0.6126</td>
<td>1.6797</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 6 Summary and Conclusion

In this thesis, we began by proposing a consistent weighted estimator to estimate the regression coefficient in a regression model using a sample from an informative sampling design. The ordinary least squares estimator (unweighted estimator) is a common choice of researchers. The probability weighted estimator is design consistent, but could have a large variance. One research goal was to find a consistent weighted estimator that is more efficient than the probability weighted estimator and that keeps the desirable properties of the probability weighted estimator. We proposed a design weighted estimator that can be more efficient than the probability weighted estimator under an informative design.

For some regression models with endogenous explanatory variables, the ordinary least squares fit can give inconsistent estimation for regression coefficients. We proposed an instrumental variable estimator that can account for sampling design and permit a test for endogeneity.

We used a test for the importance of weights in estimation to make a pretest estimator. If the null hypothesis that the design is noninformative, is accepted, we use an unweighted estimator, with the ordinary least squares estimator being a common choice. Otherwise we incorporate the sampling weights into the estimation procedure. We found the design weighted estimator and the instrumental variable estimator to perform well under complex sampling designs. Therefore, the design weighted estimator and the instrumental variable estimator can serve as good estimators in the preliminary testing procedure when the weights are determined to be necessary. We constructed a two-step preliminary testing procedure based on the test for importance of weights.
and the test for endogeneity. We concluded that the pretest estimators are compromise between unweighted estimator and weighted estimator in terms of mean squared errors.
BIBLIOGRAPHY


Sargan, J. D. (1958). The estimation of economic relationships using instrumental vari-


New York: Springer-Verlag.


Summers, R. (1965). A capital intensive approach to the small sample properties of 


Western Educational Publishing.