This is to certify that the

dissertation entitled

CORRECTING FOR SELF-SELECTION BIAS
IN CONTINGENT VALUATION

presented by

LIH-CHYUN SUN

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Agricultural Economics

Major professor
John P. Hoehn

Date 8/6/93

CORRECTING FOR SELF-SELECTION BIAS IN CONTINGENT VALUATION

By

Lih-Chyun Sun

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Economics

1993

ABSTRACT

CORRECTING FOR SELF-SELECTION BIAS IN CONTINGENT VALUATION

By

Lih-Chyun Sun

In contingent valuation (CV) studies, data can only be collected from those who are willing to participate in the studies. Results from the application of a single equation approach to this truncated sample may lead to inconsistent parameter estimates (self-selection bias). A self-selection model which contains a self-selection and a demand equation may be specified in order to detect and to correct for self-selection bias.

Based on a truncated sample, Bloom and Killingworth (1985) proposed a maximum likelihood (ML) estimator which leads to theoretically consistent parameter estimates. However, using Monte Carlo experiments, Muthén and Jöreskog (1983) showed that the estimates for parameters in the self-selection equation are not reliable even in large samples.

A self-selection model with measurement errors is proposed in this study. In the model, a CV truncated sample is transferred into a censored sample by combining survey individual data with census data which provides information for non-respondents' neighborhoods (e.g. census blocks). Based on the censored sample, two ML estimators are derived where census data are treated as if they are the true values plus errors, i.e. non-respondents' characteristics are assumed to be distributed as $N(\mu_i^*, \Sigma_i^*)$.

To apply the self-selection model with measurement errors, μ_i^* and Σ_i^* are replaced by their consistent estimates: μ_i, the average values calculated from each

census block and Σ_i, the corresponding variance-covariance matrix calculated from each census block, or Σ, the corresponding variance-covariance matrix calculated from a sample drawn from the population.

Results from Monte Carlo experiments suggest that the self-selection model with measurement errors performs well, especially when μ_i and Σ_i are adopted. The results also indicate that if the self-selection model is correctly specified, adoption of a self-selection model with measurement errors will not contaminate the original truncated sample.

The application of a self-selection model with measurement errors is not restricted to CV studies. The model can be applied to studies that adopt survey data and regression analyses.

To my parents

Dr. Chen Sun and Mrs. Feng-Chiao Lee Sun

ACKNOWLEDGEMENTS

It has been my pleasure and honor to work with my committee members, Dr. John Hoehn, Dr. Eileen van Ravenswaay, and Dr. Ching-Fan Chung. I am especially grateful to Dr. Hoehn. As my major professor, Dr. Hoehn lead me into the area of resources/environmental economics and empirical work. It was through him that I first experienced the joy of conducting research. Having worked for Dr. van Ravenswaay for the past four years, I owe her much gratitude for her guidance and tolerance. Dr. Chung enhanced my knowledge in econometrics both inside and outside the classroom. In addition to thanking him for his friendship, I thank him for introducing me to GAUSS which greatly strengthened my ability in understanding and in practicing econometrics.

I would like to express my appreciation to the Agricultural Economics Department for offering an excellent environment for studying. I am indebted to my colleague Miss Tiffany D. Phagan who spent much of her precious time editing my writing and making this dissertation readable.

Special appreciation goes to Drs. Anthony and Delia Koo. For the past seven years, Drs. Koo have been very supportive. I could never have finished my studies here at Michigan State University without their encouragement and help.

I owe a great deal to my mother-in-law, Mrs. Yu-Chueng Wu, who stayed in Lansing for a long period of time to help take care of my daughter so both my wife and I could go to school. Of course, this would never have happened if my father-in-law, Mr. I-Ming Song, was not a great gentleman.

With all my heart, I thank my parents Dr. Chen Sun and Mrs. Feng-Chiao Lee Sun for their endless love and support. Although I did not inherit their wisdom and other wonderful characteristics, I learned from them how to confront and to conquer challenges. Should I make any contribution to society, they are the persons who deserve the credit. Thanks also go to my younger brother Chih-Chyun Sun who kept my parents away from loneliness while I was abroad. I am also indebted to my aunt Diana Lee who encouraged me constantly throughout the years.

I am very fortunate to have a wonderful wife and a lovely daughter, they have sacrificed and suffered a lot to help me finish my studies. For the past years, I could have been a better father than I was. I apologize to my daughter Yihua Sun, and thank her for bringing extra happiness to the family. Lastly, with lots of love, I thank my wife Wei-Ling Song. This dissertation would never have been finished without her love, encouragement, support, and toleration.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Non-response in surveys

Contingent valuation (CV) is one of the methods used by researchers to elicit values of non-market goods. Depending on the CV survey design, the elicited values can be either a Hicksian value (i.e. compensating or equivalent variation) that is derived from a Hicksian demand function, or a consumer surplus that is derived from a Marshallian (ordinary) demand function. In many CV studies, data are collected using mail surveys. As with other survey methods, non-response is a common problem in mail surveys. The problem created by non-response is that data values intended to be observed by survey design are in fact missing. These missing values not only lead to less efficient estimates because of the reduced size of the data base, but may also lead to biased estimates due to the fact that respondents are often systematically different from non-respondents (Rubin, 1987).

In analyzing survey data, two types of possible biases can be created by non-response. The first is known as sample non-response bias, and the second is known as self-selection (or sample selection) bias (Michell and Carson, 1989). Sample non-response bias occurs when the sample distribution of some socio-economic or demographic characteristics is significantly different from the population. For example, if only low-income individuals respond to the CV surveys, the sample mean of income is then lower than the population mean of income. Sample non-response bias can be detected by comparing the sample distribution of certain socio-economic or demographic characteristics with the population distribution.

Self-selection bias occurs when the non-response is non-random, which means that the reasons for non-response are endogenous to the survey study. For example, only those who have a higher marginal propensity to consume the non-market good respond to the CV survey. Unlike sample non-response bias, it is difficult to find a simple indicator for detecting the existence of self-selection bias.

Non-response can usually be divided into two categories, namely, item non-response and unit non-response. In CV mail surveys, item non-response means that a respondent returns the survey but fails to answer some of the questions; unit non-response indicates that a member of the sample fails to return the survey. Both item and unit non-response can cause either sample non-response or self-selection bias, or both.

One way to compensate for item non-response is to replace those missing values with imputed values (Little and Rubin, 1987, Rubin, 1987). An alternative is to use a generalized Heckman's two-stage method to correct for the possible biases that are caused by the item non-response (Ong et.al., 1988). However, item non-response is not the concern of this study and statistical methods that are related to item non-response will not be discussed here.

The purposes of this study are first to distinguish the differences between sample non-response and self-selection biases and then to develop parametric analyses to detect and to correct for the possible self-selection bias that is caused by unit non-response.

1.2 Self-selection and sample non-response biases

In order to derive values of non-market goods in CV studies, a demand (or inverse demand) function is estimated by regression analyses.[1] In this section, self-selection and sample non-response biases are first examined under a regression framework. Next, graphs based on simplified models are provided to demonstrate intuitively the relationships between self-selection and sample non-response biases.

1.2.1 Self-selection and sample non-response biases: a regression analysis

Suppose that individual i's demand for a non-market good, Y, is described by a linear structural equation

$$y_i = x_i'\beta + u_i,\ i = 1, 2, ..., N,$$

where x_i is a column vector of stochastic variables, u_i is an error term, and $E(u_i \mid x_i) = 0$ (i.e. $E(y_i \mid x_i) = x_i'\beta$). Suppose now that the resulting OLS regression using only the data from respondents is

$$y_i = x_i'\theta + e_i,\ i = 1, 2, ..., M, \text{ and } M < N.$$

---

[1] A total value function that determines an individual's willingness to pay (WTP) for a perceived environmental change is usually estimated in a CV study (Randall, 1987, p.260). The relationship between WTP's and perceived environmental changes can be thought of as a demand (inverse demand) function. In some studies, researchers use CV to estimate the demand for a market good with non-market attributes (van Ravenswaay and Hoehn, 1991a, 1991b). For convenience, the following analyses concentrate only on a demand function.

By definition, self-selection bias occurs as a result of

$$E(e_i \mid x_i, \text{return the survey}) \neq 0,$$

and results of self-selection bias are

$$E(\theta \mid x_i) \neq \beta, \text{ and}$$

$$x_i'\theta \neq E(y_i \mid x_i) = x_i'\beta.$$

Sample non-response bias was previously defined as the sample distribution of x_i differing significantly from the population distribution. For example, if the sample mean of x ($= [x_1\ x_2\ ...\ x_M]'$) is significantly different from the population mean x^*, it is then suspected that the sample mean of y ($= [y_1\ y_2\ ...\ y_M]'$) is different from the population mean y^*. Based on neoclassical regression analyses (Goldberger, 1991, Chapter 25), however, $E(\theta \mid x_i) = \beta$ is always true given that there is no self-selection bias. Given the population mean x^*, y^* is simply the conditional expectation $E(y \mid x = x^*)$, and can be calculated consistently by $x^{*\prime}\theta$. Apparently, the distribution of x_i does not play a role in regression models.

In CV studies, for a demand function derived by regression analyses, $E(y_i \mid x_i) = x_i'\beta$ is always true provided that there is no self-selection bias, i.e. $E(e_i \mid x_i$, return the survey) $= 0$. This holds regardless of any difference between x_i and population distributions. In other words, in regression analyses, sample non-response bias does not affect the consistency of the parameter estimates. As long as the parameter estimates are consistent, the conditional expectation of y_i given x_i can always be calculated consistently. Rather than worrying about sample non-response bias, researchers should instead focus their attention on self-selection bias.

In analyzing CV survey data, self-selection and sample non-response bias are two very different issues; there exists no special relationship between them.

Self-selection causes biased (inconsistent) parameter estimates due to the non-zero conditional expectation of the error term given the independent variables. Sample non-response bias does not even play a role in regression analyses.

1.2.2 Self-selection and sample non-response biases: a graphical analysis

It is helpful to demonstrate graphically the possible relationships between self-selection and sample non-response biases in the following examples.

Suppose that the relationship between an individual's demand for a non-market good, Y, and his/her income is

$$y_i = c + x_i\beta + u_i,$$

where c is a constant term and x_i is assumed to be individual i's income, and the OLS regression using data from only the respondents is

$$y_i = d + x_i\theta + e_i.$$

Further, assume that the population (sample) mean for y and for income are y^* ($\overline{y}$) and x^* ($\overline{x}$) respectively.

In the following graphs, a solid line represents the sample regression line, a broken line represents the population regression line, and the marginal propensity to consume the non-market good is defined as dy_i/dx_i.

Example 1. Presence of sample non-response bias, absence of self-selection bias (Figure 1.1).

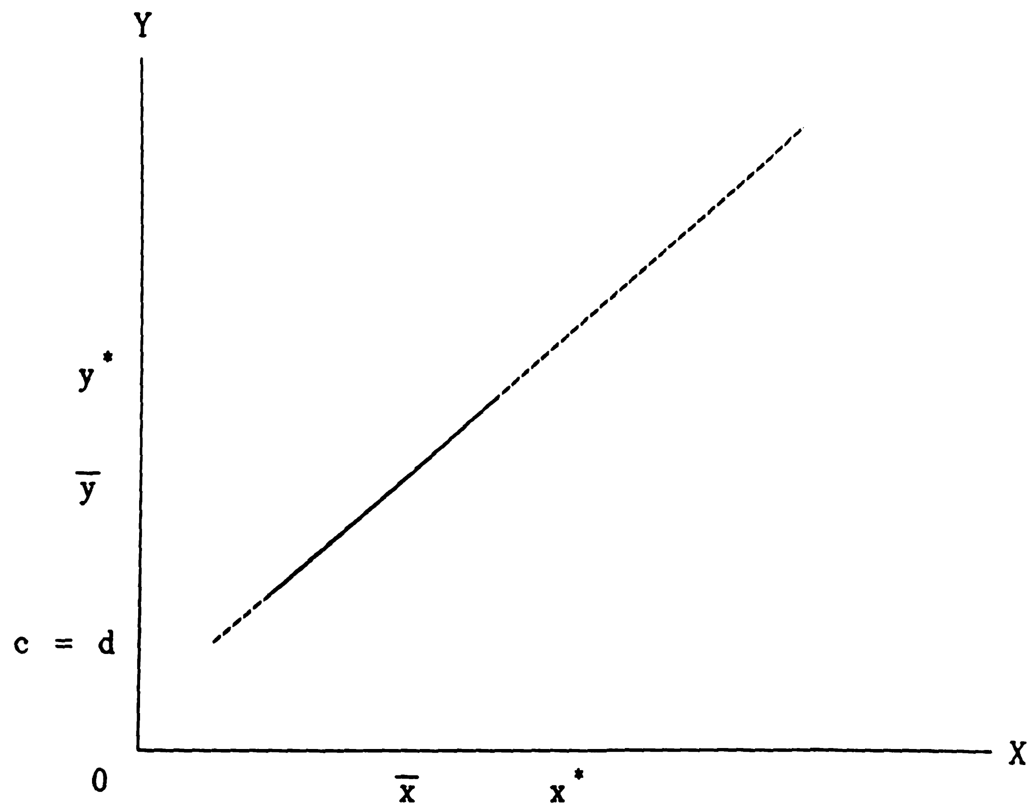

Figure 1.1 Presence of sample non-response bias, absence of self-selection bias

Only low-income individuals return the surveys, and both low- and high-income individuals have the same marginal propensity to consume the non-market good, i.e.

$$y^* > \bar{y},\ x^* > \bar{x},$$

$$d = c, \text{ and } \theta = \beta.$$

Example 2. Presence of both self-selection and sample non-response biases (Figure 1.2).

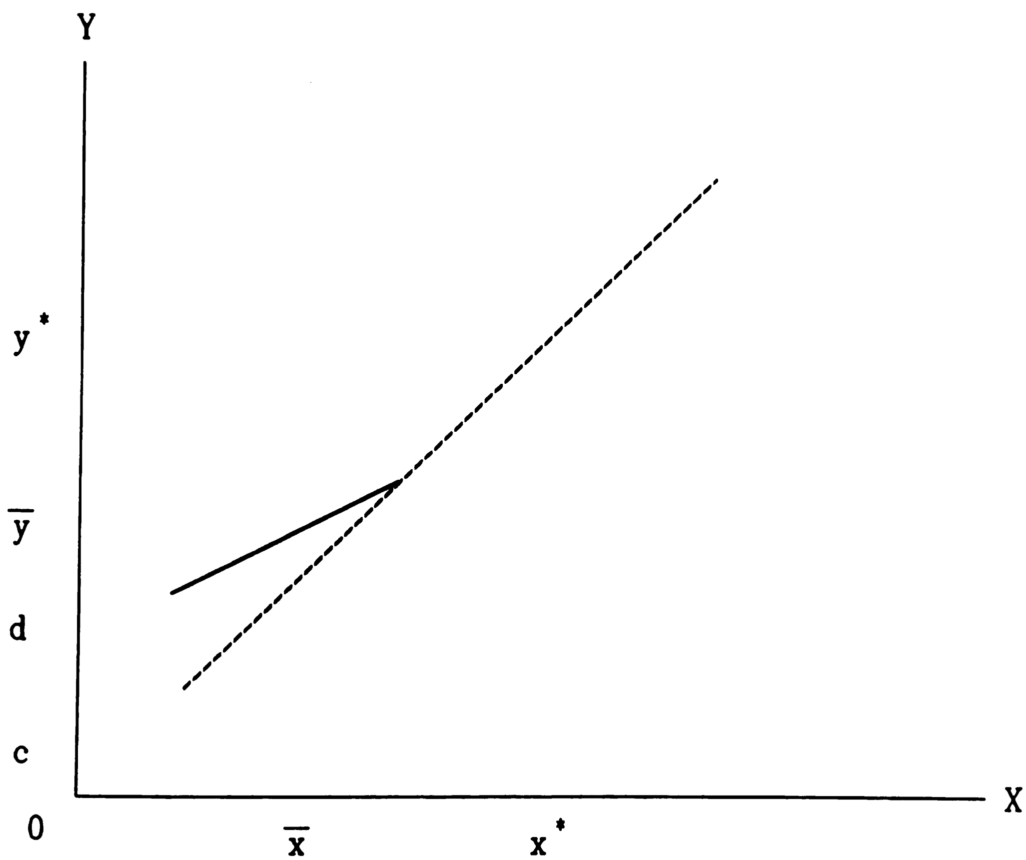

Figure 1.2 Presence of both self-selection and sample non-response biases

Only some of the low-income individuals return the surveys, and those low-income individuals have a lower marginal propensity to consume the non-market good than do other individuals, i.e.

$$y^* > \bar{y},\ x^* > \bar{x},$$

$$d \neq c, \text{ and } \theta \neq \beta.$$

Example 3. Presence of self-selection bias, absence of sample non-response bias (Figure 1.3).

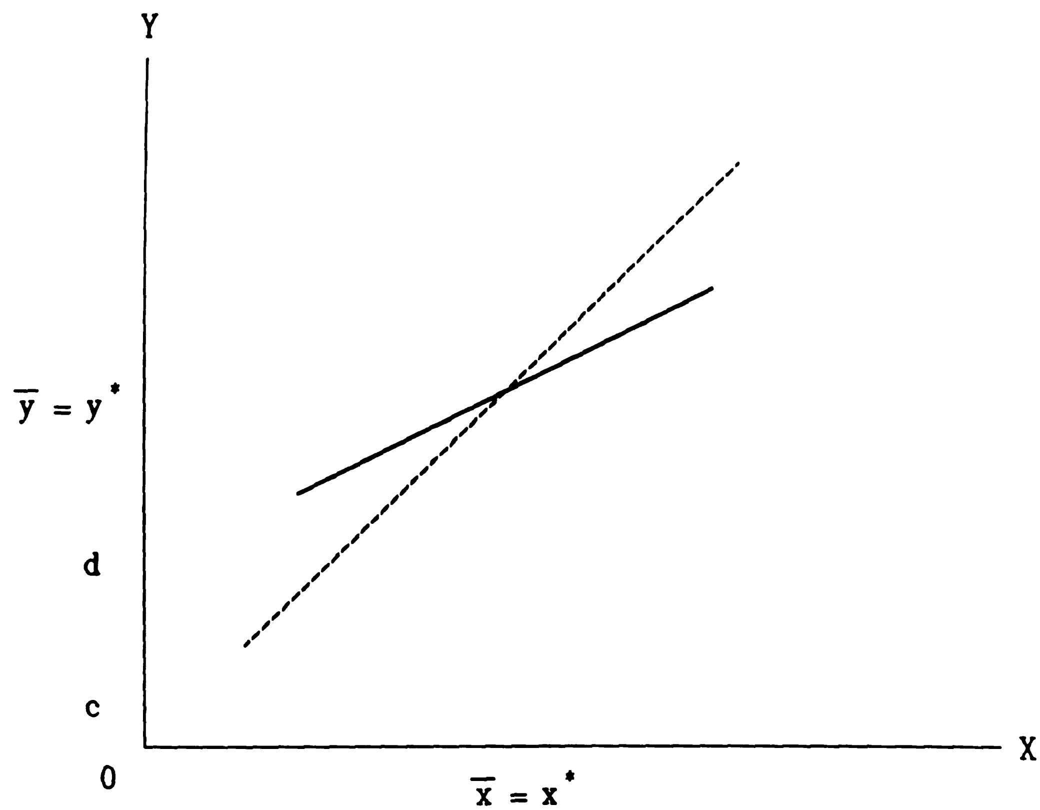

Figure 1.3 Presence of self-selection bias, absence of sample non-response bias

Only those who have a lower marginal propensity to consume the non-market good return the surveys, i.e.

$$y^* = \overline{y},\ x^* = \overline{x}, \text{ but}$$

$$d \neq c, \text{ and } \theta \neq \beta.$$

Example 4. One special case is when self-selection bias affects only the estimate of the constant term (Figure 1.4).

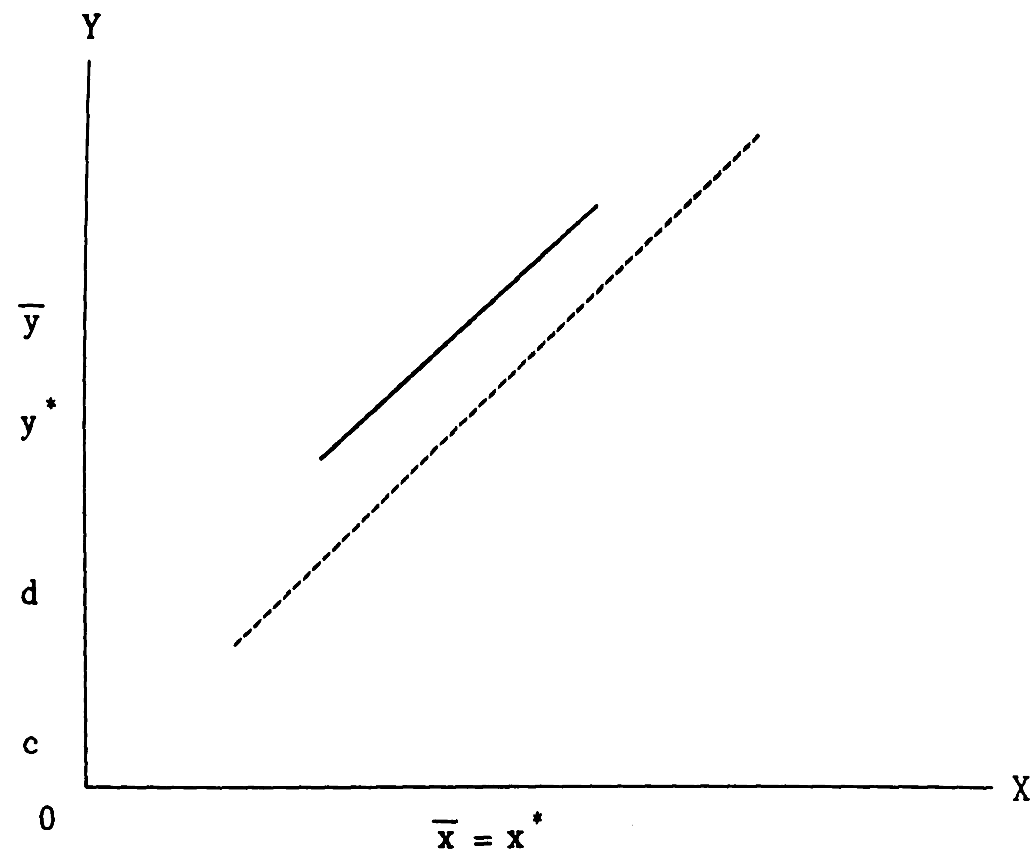

Figure 1.4 Self-selection bias affects only the constant term

Both low- and high-income individuals have the same marginal propensity to consume the non-market good, but the average consumptions are different, i.e.

$$d \neq c, \text{ but } \theta = \beta.$$

Given the above examples, there is no special relationship between self-selection and sample non-response biases that can be observed. Thus, sample non-response bias is not an appropriate indicator for the presence of self-selection bias. In CV studies, if regression analyses are adopted, examining the presence of sample non-response bias will not help researchers to detect the presence of self-selection bias. Furthermore, the consistency of parameter estimates has nothing to do with the sample non-response bias.

1.3 Self-selection in contingent valuation

Although self-selection bias has been considered as a potential problem in CV studies (Mitchell and Carson, 1989; Edwards and Anderson, 1987; Loomis, 1987), none of the existing studies has demonstrated any empirical evidence regarding self-selection bias in CV studies. However, some studies indicate that sample non-response bias may be a potential problem when surveys are used for collecting data. For example, several studies have investigated the factors which affect individuals' decisions to answer surveys (Green, 1991; Green and Kvidahl, 1989; Green and Stager, 1986; Goyder, 1982; Brown, et al. 1981; Kanuk and Berenson, 1975). It is pointed out that respondents tend to be older, have higher income and higher education. In some cases, respondents and non-respondents have different characteristics with respect to occupation, residence location, and gender. In some studies, it has been found that respondents are more interested than non-respondents in the topic of the survey studies. For example, Whitehead (1991) showed that members of environmental interest groups responding to CV surveys that value environmental goods have a particular interest in the topic of the survey. Brown et. al. (1989) found that mail response rates were higher among members of environmental interest groups. Walsh, et. al. (1984), and

Bowker and Stoll (1988) have suggested that members of environmental interest groups hold larger environmental values than non-members when measured in CV markets.[2]

As previously stated, CV is used to derive values of non-market goods through an estimated demand function. In economic modeling, certain socio-economic and demographic characteristics are frequently used as the explanatory variables in estimating a demand function. These explanatory variables may include income, age, education, gender, and location. These same variables have also been considered as factors affecting an individual's response decision.

Three observations can be drawn from the studies cited above. First, certain common variables explain both individuals' decisions to answer the survey and their demand for non-market goods. Second, respondents and non-respondents may possess different characteristics. Finally, respondents may be more interested in the survey topics than non-respondents.

Although these observations do not offer direct evidence for the presence of self-selection bias, they do call attention to the potential existence of self-selection bias. For instance, examples 2, 3, and 4 (Figures 1.2, 1.3, and 1.4, Section 1.2.2) raised in the previous section gave several possible outcomes that indicated the co-existence of sample non-response and self-selection biases.

Since sample non-response bias alone is not a sufficient indicator of the presence of self-selection bias, and since none of the existing studies has provided satisfactory work to detect and to correct for self-selection bias empirically, self-selection bias remains an empirical hypothesis that should be tested.

---

[2] In estimating recreation demand, several authors have noticed the problem of self-selection bias when data is collected from on-site and user groups (Shaw, 1988; Smith, 1988; Bockstael, et. al., 1990). However, this type of self-selection bias is caused by sampling method.

1.4 Plan of work

This study concentrates on parametric analyses in testing and in correcting for potential self-selection bias in CV studies. Respondents' behavior is modeled as a two-step decision making process. The first step is concerned with a respondent's decision whether to return the survey or not. Under a random utility framework, it is assumed that an individual gains utility from answering the survey. The cost of answering the survey is the opportunity cost of the time that is required for answering the CV survey. Based on utility maximization subject to both budget and time constraints, an individual will answer and return the survey only if the net utility gain is positive.[3] The net utility gain can be summarized by an equation referred to here as a self-selection equation. If the individual decides to return the survey, the second step is to determine the respondent's demand for the non-market good through a demand function. The complete self-selection model is described by estimating the self-selection and demand equations simultaneously using maximum likelihood (ML) estimators.

Three types of ML estimators are considered in this study. The first ML estimator is used when the sample is truncated. In a truncated sample, neither non-respondents' characteristics nor their demand for the non-market goods can be observed by researchers. The second ML estimator is used when the sample is censored. With a censored sample, non-respondents' characteristics are observable, but demand for the non-market goods is not observed.

The third ML estimator transforms a truncated sample into a censored sample by adopting information from both the truncated sample and census data. For example, in mail surveys, although researchers do not observe anything from

[3] In this study, respondents who returned the surveys with incomplete answers are treated as non-respondents.

non-respondents, mailing addresses are usually available, and average characteristics of non-respondents' neighborhoods can be acquired from census data. If a non-respondent's characteristics can be treated as the average characteristics of his/her neighborhood plus an error, this can provide researchers with additional information that can be used in regression analyses.

Two types of questionnaires are frequently used in CV studies. The first type is the open-ended questionnaire and the second type is the closed-ended questionnaire. For an open-ended questionnaire, demand is observed as a continuous variable (or is sometimes censored at certain values). For a closed-ended questionnaire, demand is not directly observable. Given a referendum price, only a YES/NO answer is observed. In this study, data from both open- and closed-ended questionnaires are discussed along with the three different ML estimators.

The remaining chapters of this study are completed in the following manner. Chapter 2 describes the statistical nature of self-selection bias and the rationale of testing and correcting for self-selection bias. Studies concerned with self-selection bias under both truncated and censored samples are also reviewed. Chapter 3 derives a self-selection model under the random utility framework. In addition, ML estimators that transform a truncated survey sample into a censored sample by adopting information from both the truncated sample and census data are also developed. The resulting estimates from the ML estimators are examined by Monte Carlo experiments, and the results are summarized in Chapter 4. Finally, concluding remarks are given in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Self-selection has been considered by economists, particularly so by labor economists, for some time. In most of the studies that include this issue, self-selection is used in modeling the earnings among different sectors. For example, let y_1 and y_2 represent the potential earnings for two sectors. An individual will work in sector 1 only if $y_1 > y_2$, then $E(y_1 \mid y_1 > y_2) \neq E(y_1)$. In a conventional self-selection model, two separate equations are used to model y_1 and y_2. Based on income maximization, a latent variable is modeled by a third equation which describes I^* as a function of $(y_1 - y_2)$. The individual will work in sector 1 if $I^* > 0$ $(y_1 > y_2)$, and in sector 2 otherwise. An econometric model that is designed for this type of self-selection is often called a switching regression model.

Willis and Rosen (1979) modeled the demand for college attendance based on the comparative advantage in expected lifetime earnings. Considering simultaneously the demand for and supply of labor, Heckman and Sedlacek (1985) presented a model of the sectoral allocation of workers from different demographic types. What made their study unique is their use of aggregate data to predict earnings for the different sectors, combining these predicted earnings and micro data to estimate the labor supply in the different sectors. In his study, Borjas (1987) modeled the earnings for immigrants based on the difference in wages earned in the U. S. and potential wages in their native countries. Recently, Heckman and Sedlacek (1990) modeled self-selection based on utility maximization, instead of self-selection based on earnings.

In a CV study, a self-selection hypothesis would assert that only those who have enough interest in the topic of study will return the surveys, and that respondents have different demand behavior than non-respondents. Under this hypothesis, self-selection in a CV study differs from self-selection in a conventional switching regression model in two ways. Borrowing the above two sector earnings model, assume y_{i1} and y_{i2} are individual i's demand for goods 1 and 2 respectively. First, instead of modeling demand for both goods 1 and 2, a CV study usually models only the demand for one good (say, y_{i1}). Second, the self-selection criterion in a CV study is the net utility gain from answering the survey, while the criterion in a conventional switching regression model is the potential difference in demand (y_{i1} - y_{i2}).

Since only the demand for one good (y_{i1}) is modeled in a CV study, the self-selection model is less complicated than a conventional switching regression model. With this simplification, the nature of switching regression models is left intact but the statistical process for estimating the model is simplified.

2.2 A self-selection model

Under a CV framework, consider the following self-selection model regarding individual i's demand for a good Q (q_i). Individual *i*'s *self-selection* behavior is governed by the self-selection equation

$$I_i^* = z_i'\gamma + u_i, \; i = 1,2,...,N,$$

$$I_i = 1 \,, \quad \text{iff} \;\; I_i^* > 0,$$

$$I_i = 0 \,, \quad \text{otherwise.}$$

In the self-selection equation, I_i^* is unobservable, but I_i is observed. If $I_i = 1$, the demand equation,

$$q_i = x_i'\beta + e_i,$$

is then observed. In the self-selection and demand equations, z_i (x_i) is a kx1 (mx1) vector of exogenous variables, γ (β) is a kx1 (mx1) vector of parameters to be estimated, u_i (e_i) is a random error. Assume that

$$u_i \sim \text{i.i.d. } N(0,\ 1),$$

$$e_i \sim \text{i.i.d. } N(0,\ \sigma^2),$$

$$(e_i,\ u_i) \sim \text{i.i.d. } BN(0,\ 0,\ \Omega),$$

where N(.,.) and BN(.,.,.) are a univariate and a bivariate normal distribution respectively, and

$$\Omega = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

is the variance-covariance matrix.

Suppose that researchers are interested in estimating the demand equation. Since q_i is observed only if $I_i = 1$, the distribution of q_i is truncated, and the demand equation alone does not correctly specify the demand for the good (q_i). To specify the demand equation correctly, the endogeneity that is caused by the self-selection behavior must be taken into account. Thus, a correct model specification is described jointly by the self-selection and the demand equations, and the objective is to obtain consistent estimates for β, γ, ρ, and σ^2.

Under a CV framework, I_i^* in the self-selection equation can be thought of as individual i's net utility gain from answering the survey. Self-selection implies that an individual's decision to answer and to return the survey is correlated with the topic of study (i.e. q_{li}, demand for the good Q). In other words, the decision to answer and to return the survey is endogenous to the study.

2.3 Estimators

In this section, three estimators for correcting self-selection bias are reviewed. Heckman's two-stage estimator is first examined,[1] followed by two ML estimators that are based on either a censored or on a truncated sample.

2.3.1 Heckman's two-stage estimator

Based on the moments of a truncated bivariate normal distribution, according to the self-selection model described by the self-selection and demand equations, Heckman (1979, 1976) demonstrated that

[1] Although Heckman's two-stage estimator is well known, it provides a clear and straightforward explanation of the nature of self-selection bias.

$$
\begin{aligned}
&E(q_i \mid x_i, I_i = 1) \\
&= x_i'\beta + E(e_i \mid I_i = 1) \\
&= x_i'\beta + E(e_i \mid u_i > -z_i'\gamma) \\
&= x_i'\beta + \alpha \frac{\phi(-z_i'\gamma)}{1 - \Phi(-z_i'\gamma)} \\
&= x_i'\beta + \alpha \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)},
\end{aligned}
$$

where $\alpha = \rho\ \sigma$ (i.e. the covariance between e and u), ϕ and Φ are standard normal density and distribution functions respectively, and $\phi(\cdot)/\Phi(\cdot)$ is the inverse Mill's ratio, or in some contexts, the hazard rate. The equation $E(q_i \mid x_i, I_i = 1)$ clearly indicates that OLS regression of q on x leads to inconsistent estimates for β.[2]

Based on the equation $E(q_i \mid x_i, I_i = 1)$, Heckman proposed a two-stage method for estimating β, γ, and α. In the first stage, according to the self-selection equation, a probit model is used to obtain $\hat{\gamma}$, a consistent estimate of γ. The predicted inverse Mill's ratio is then calculated as $\phi(z_i'\hat{\gamma})/\Phi(z_i'\hat{\gamma})$. In the second stage, using only the returned surveys, consistent estimates of β and α can be obtained by regressing q on x and the predicted inverse Mill's ratio using standard OLS procedures.[3] Self-selection bias can then be detected by testing the hypothesis that $\alpha = 0$ (i.e. $\rho = 0$).

---

[2] Unless $\phi(\cdot)/\Phi(\cdot)$ is orthogonal to x, there is omitted variable bias.

[3] Due to heteroscedasticity, feasible generalized least squared (FGLS) can also be used (Greene, 1990, pp. 739 - 747). FGLS may result in more efficient estimates.

Based on the equation $E(q_i \mid x_i, I_i = 1)$, self-selection bias can be viewed as an omitted variable bias, where the omitted variable is the inverse Mill's ratio.

Several observations concerning self-selection bias can be drawn from the equation $E(q_i \mid x_i, I_i = 1)$. First, there is no self-selection bias if e_i and u_i are uncorrelated ($\rho = 0$). Second, since $|\rho| \in [0, 1]$, if σ is small, and Φ is close to 1, then $\rho \cdot \sigma \cdot [\phi(\cdot)/\Phi(\cdot)]$ can be very close to zero. If this is the case, there will be no self-selection bias. This can happen when the response rate is high (i.e. for each individual, $\phi(\cdot)/\Phi(\cdot)$ is close to 0). Third, as pointed out by Heckman (1976), if $\rho \cdot \sigma \cdot [\phi(\cdot)/\Phi(\cdot)]$ is a constant, then all of the slope coefficients estimated using OLS (where only the returned surveys are used) are consistent except for the constant term[4] (this can also be seen from Figure 1.4, Chapter 1, Section 1.2.2). Fourth, $E(q_i \mid x_i, z_i)$, the conditional expectation given x_i and z_i (unconditional on whether the survey is returned or not), can be derived as

$$
\begin{aligned}
& E\,(q_i \mid x_i, z_i) \\
&= \left[\, x_i'\beta + \alpha \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)} \,\right] \cdot \Phi(z_i'\gamma) + \\
&\quad \left[\, x_i'\beta - \alpha \frac{\phi(z_i'\gamma)}{\Phi(-z_i'\gamma)} \,\right] \cdot \Phi(-z_i'\gamma) \\
&= x_i'\beta \cdot [\Phi(z_i'\gamma) + \Phi(-z_i'\gamma)] \\
&= x_i'\beta.
\end{aligned}
$$

Although Heckman's two-stage method offers a convenient way to estimate and to test for self-selection bias, it suffers from three major disadvantages. First, in order to estimate γ and to calculate the inverse Mill's ratio, the z matrix for

---

[4] This could happen if $\gamma = 0$. Recall that $E(q_i \mid x_i, I_i = 1) = x_i'\beta + E(e_i \mid u_i > -z_i'\gamma)$, if $\gamma = 0$, then $E(e_i \mid u_i > 0) = \rho\ \eta$, which is a constant.

both respondents and non-respondents must be known. In a truncated sample, z is not known for non-respondents. Hence, the Heckman's two-stage method cannot be applied to data from a truncated sample. Second, since only the returned surveys are used in the second stage, it is less efficient than if the full sample is used.[5] Third, the conventional formula used in OLS to calculate the variance-covariance matrix does not provide the correct variance-covariance matrix for the second stage OLS estimation.

Instead of using Heckman's two-stage method, one alternative for deriving consistent estimates is to estimate the self-selection and the demand equations jointly by ML estimators. Depending on the nature of the sample, there are essentially two types of likelihood functions that can be specified. The first type of likelihood function is specified for a censored sample and the second type for a truncated sample.

2.3.2 Self-selection with a censored sample

According to the self-selection and the demand equations, a censored sample indicates that for those whose $I_i = 0$ (individual i did not return the survey), x_i and z_i can still be observed. In a censored sample, for those who did not return the survey, z_i is still available and can be used to explain individual i's self-selection behavior. In addition, the explanatory variables, x_i, in the demand equation can also be observed. For a censored sample, consistent and efficient estimates for β, γ, ρ, and σ^2 can all be obtained by maximizing the likelihood function[6]

---

[5] It is also possible to use the full sample (Maddala, 1983, p. 159).

[6] In practice, it is the log-likelihood function that is maximized. However, the likelihood function simplifies interpretation.

$$L_c = \prod_{I_i=1} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du$$

$$\prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-z_i'\gamma} g(e,\ u,\ \Omega)\ du\ de,$$

where g(.,.,.) is a bivariate normal density function. In the likelihood function, L_c, the first term is the likelihood for those who returned the survey and is the product of the conditional density of q_i given that individual i returned the survey. The second term is the likelihood for those who did not return the survey and is the product of the joint distribution function.[7]

2.3.3 Self-selection with a truncated sample

By definition, a censored sample implies that all non-respondents' x's and z's can still be observed. In practice, this does not seem to be the case for most of the CV studies. Very often, data used in CV studies are truncated; namely, neither x's nor z's can be observed from non-respondents.

According to the self-selection and the demand equations, a truncated sample indicates that when $I_i = 0$, all q_i, x_i, and z_i are not observed. In other words, we know nothing of those who did not return the survey. In the case of a truncated sample, the self-selection and demand equations can be estimated jointly by maximizing the likelihood function

[7] For further discussion of self-selection models under a censored sample, see Little (1985), Lee (1984), Goldberger (1981), Greene (1981), Olsen (1980), and Nelson (1977).

$$L_t = \prod_{I_i=1} \frac{\int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du}{\Phi(z_i'\gamma)},$$

and the estimated β, γ, σ^2, and ρ are consistent (Bloom and Killingsworth, 1985).[8] The likelihood function, L_t, is based on a truncated normal distribution. The numerator of i's likelihood is the conditional density of q_i given that individual i returned the survey, and the denominator is the probability that individual i returns the survey. In a self-selection model with a truncated sample, only the information from returned surveys is available for use in estimation.[9]

In most CV studies, data used in econometric analyses are obtained from truncated samples. According to Bloom and Killingsworth (1985), self-selection models with a truncated sample should not create problems in econometric analyses when the ML estimator is applied. However, bearing in mind that x_i and z_i are likely to have variables in common, or at least to be highly correlated, it seems unlikely that one would be able to obtain good estimates of parameters other than β (Pudney, 1989, p. 83). A study by Muthén and Jöreskog (1983) using Monte Carlo experiments tends to confirm this suspicion and shows that the estimate for γ is not reliable even in large samples,[10] although it is possible to correct for self-selection bias in the β coefficients. This is a major disadvantage of using data from a truncated sample.

---

[8] The same likelihood function is also presented by Maddala (1983, pp. 150 and 267).

[9] Unlike the studies conducted by Hauseman and Wise (1981, 1977) where the data are acquired from a sample that is truncated by an exogenous variable. A truncated sample in this study is a sample that is truncated by an endogenous variable.

[10] Part of Muthén and Jöreskog's (1983) results are reported in Appendix A.

2.4 Summary

The self-selection model considered in this study consists of two components. The first is the self-selection equation which is a probit-type equation, and the second is a demand equation. Under a CV framework, the analyses began with an examination of self-selection models with either a censored or a truncated sample. Econometric analyses with a censored sample were found to have preferred properties (i.e. consistency and efficiency). However, censored samples are generally not available for most CV studies. The majority of CV studies use surveys to collect data. Since data is collected only from respondents, the sample is truncated. According to existing econometric methods, in order to test and to correct for self-selection bias in CV studies, the ML estimator that is based on a truncated sample (Bloom and Killingsworth, 1985) must be adopted. In theory, ML estimators lead to consistent and efficient estimates provided that the likelihood function is correctly specified. However, Monte Carlo experiments have shown that in truncated samples, parameters in the self-selection equation could not be estimated reliably even with large samples (Muthén and Jöreskog, 1983).

This disadvantage of the ML estimator that is based on a truncated samples motivates the derivation of ML estimators in the next chapter which combine individual survey data with census data and transfer a truncated sample into a censored sample.

CHAPTER 3

SELF-SELECTION MODELS WITH MEASUREMENT ERRORS

A self-selection model consists of two correlated components. The first component is a self-selection equation which is essentially a probit model. The second component is a demand equation. Under self-selection, an individual's demand is observed only if the corresponding latent variable which is generated by the self-selection equation has a value greater than zero. In this study, the self-selection models differ from conventional models. For those individuals with unobserved demand, all of the independent variables in the self-selection equation are observed but with measurement errors.

This chapter begins by describing a self-selection equation. The self-selection equation is developed using the random utility model under a CV framework. A probit model with measurement errors is derived where the r.h.s. variables are measured with errors whenever the l.h.s. latent variable has a value less than or equal to zero. A self-selection model with measurement errors is developed based on the probit model with measurement errors described above and a linear demand equation. Finally, the model is generalized to allow for qualitative and limited dependent variables in the demand equation.

3.1 Self-selection based on a random utility model under a CV framework

Assume a CV study uses mail surveys to elicit the demand for a good Q. In addition, the questionnaires used in the survey are open-ended. Individual i's demand for the good is q_i.[1] However, q_i is observed only if individual i returned the survey and gave valid answers.[2]

Suppose that individual i's decision to return the survey is based on his net utility gain from answering the survey. Individual i will return the survey only if the net utility gain from answering the survey is positive. However, individual i's net utility gain cannot be observed directly; only the realization of the net utility gain (i.e. to return or not to return the survey) is observed.

To model an individual's self-selection behavior, assume that an individual maximizes utility subject to both a budget and a time constraint:

$$\begin{aligned} &\text{Max. } U(C, L, t\cdot I, I \mid s) \\ &\text{s.t. } w\cdot T^{*} - w\cdot L - w\cdot t\cdot I = P\cdot C, \\ &T^{*} = T + L + t\cdot I, \end{aligned} \tag{1}$$

where C is a composite good that the individual consumes at price P, L is leisure time spent, t is the time devoted to answering the survey, I is an indicator which equals 1 if he answers the survey and 0 otherwise, w is the wage rate, s is a vector of socio-economic and demographic variables other than the wage rate, T^* is the total time available (which is fixed), T is the time devoted to market work and is also assumed to be fixed. At maximum utility,

$$U_L = \frac{U_I}{t}, \tag{2}$$

where U_L and U_I are the marginal utilities of leisure and of answering the survey respectively. At utility maximization, equation (2) states that the marginal utility

---

[1] At this stage, q_i is assumed to be continuous and $-\infty < q_i < \infty$. This assumption will be released later in this chapter.

[2] All of the returned surveys are assumed to have valid answers. This assumption excludes the case of item non-response.

per unit of time for answering the survey equals the marginal utility of leisure (i.e. the marginal utility of answering the survey is equal to the marginal utility of leisure multiplied by the time used in answering the survey).

The individual's indirect utility function can be written as

$$U^{*}(P,\ w,\ T,\ t \mid s,\ I). \tag{3}$$

Let C be the numeraire, and set P = 1. Furthermore, assume that t is constant across individuals. Since T is treated as fixed, the indirect utility function becomes

$$U^{**}(Y \mid s,\ I), \tag{4}$$

where Y (= w·T) is the individual's income. The condition for an individual to answer and to return the survey is

$$\begin{aligned} &U^{**}(Y \mid s,\ I = 1) - U^{**}(Y \mid s,\ I = 0) \\ &= V(Y,\ s) \\ &= V(z) > 0, \end{aligned} \tag{5}$$

where z = (Y, s) is a vector of socio-economic and demographic variables (including income).

Assume V(z) is a linear function of all elements in z, and u is a random error drawn from a standard normal distribution. Individual i's self-selection equation can be expressed by a standard probit model:

$$I_i^* = z_i'\gamma + u_i, \; u_i \sim \text{i.i.d. } N(0,\, 1),$$

$$I_i = 1, \text{ iff } I_i^* > 0, \tag{6}$$

$$I_i = 0, \text{ otherwise.}$$

Equation (6) is the self-selection equation that models individual i's decision behavior. In equation (6), I_i^* is the net (indirect) utility gain and cannot be observed, z_i is a column vector consisting of exogenous variables (including income) that explain individual i's net (indirect) utility gain, γ is a column vector of parameters to be estimated, and N(0, 1) represents a standard normal distribution. Although I_i^* cannot be observed, researchers can observe I_i.

3.2 A probit model with measurement errors

In this study, measurement errors occur when proxy variables are used to approximate the true values of the exogenous variables in the self-selection equation for non-respondents.

As stated earlier, the self-selection equation is essentially a probit model. Before the self-selection model with measurement errors can be studied, a probit model with measurement errors must be discussed.

3.2.1 Derivation of the probit model with measurement errors

Following the notation used in the previous sections, the derivation of a probit model with measurement errors begins with

$$I_i^* = z_i'\gamma + u_i,\ u_i \sim \text{i.i.d. } N(0,\ 1). \tag{7}$$

As before, I_i^* is a latent variable, z_i is a kx1 vector of independent variables, γ is a kx1 vector of parameters, and u_i is an error term drawn from a standard normal distribution. I_i^* cannot be observed, however, I_i can be observed. In addition, I_i equals 1 if $I_i^* = z_i'\gamma + u_i > 0$; 0, otherwise.

For a respondent, $I_i = 1$, z_i can be observed, and the likelihood for the respondent is derived as:

$$\begin{aligned} & z_i'\gamma + u_i > 0 \\ & \rightarrow u_i > -z_i'\gamma \\ & \rightarrow \text{Prob}(u_i > -z_i'\gamma) \\ & \quad = \left[1 - \Phi\left(-z_i'\gamma\right)\right] = \Phi\left(z_i'\gamma\right), \end{aligned} \tag{8}$$

where $\Phi(\cdot)$ is a standard normal distribution function.

For a non-respondent, $I_i = 0$, z_i cannot be observed. However, μ_i, which is the average value of z_i, is estimated using a random sample drawn from individual i's neighborhood (e.g. a census block[3]).[4] Let n_i be the size of the random sample and $z_i \sim N(\mu_i^*, \Sigma_i^*)$. Obviously,

$$\mu_i \sim N(\mu_i^*, \frac{\Sigma_i^*}{n_i}). \tag{9}$$

---

[3] This can be a census block, a county, a state, or even a region. For convenience, a census block is used in the following analyses.

[4] For example, this can be done by matching the mailing list with the census block to obtain the average value of each z_i available in the census data.

Define measurement errors as $v_i = z_i - \mu_i$, then

$$v_i \sim N\left(0, \frac{\Sigma_i^*}{n_i} + \Sigma_i^* - 2\frac{\Sigma_i^*}{n_i}\right)$$
$$\rightarrow v_i \sim N\left(0, \frac{n_i - 1}{n_i}\Sigma_i^*\right). \tag{10}$$

In general, $\frac{(n_i - 1)}{n_i} \approx 1$, so the distribution of v_i can be approximated by[5]

$$v_i \sim N(0, \Sigma_i^*). \tag{11}$$

To derive the likelihood for a non-respondent, the self-selection equation for a non-respondent can be written as

$$z_i'\gamma + u_i \leq 0$$
$$\rightarrow (\mu_i + v_i)'\gamma + u_i \leq 0 \tag{12}$$
$$\rightarrow \mu_i'\gamma + w_i \leq 0, \text{ and } w_i = u_i + v_i'\gamma.$$

Further, assume that u_i and v_i are independent. Then

$$w_i \sim N(0, \varphi_i^2), \text{ and}$$
$$\varphi_i^2 = 1 + \gamma'\Sigma_i^*\gamma. \tag{13}$$

---

[5] Alternatively, the unobserved z_i can be decomposed into the sum of a deterministic component, μ_i^*, and a random component, v_i, with $v_i \sim N(0, \Sigma_i^*)$. Now replace μ_i^* with its consistent estimate, μ_i. We have $z_i = \mu_i + v_i$, and $v_i \sim N(0, \Sigma_i^*)$.

The likelihood for a non-respondent can then be derived as[6]

$$
\begin{aligned}
& \mu_i'\gamma + w_i \le 0,\ w_i \sim N(0,\ \varphi_i^2) \\
& \rightarrow w_i \le -\mu_i'\gamma \\
& \rightarrow \mathrm{Prob}(w_i \le -\mu_i'\gamma) \\
& = \Phi\left(\frac{-\mu_i'\gamma}{\varphi_i}\right).
\end{aligned}
\tag{14}
$$

Based on equations (8) through (14), the likelihood function for the probit model with measurement errors is

$$
L_S = \prod_{I_i=1} \Phi\left(z_i'\gamma\right) \cdot \prod_{I_i=0} \Phi\left(\frac{-\mu_i'\gamma}{\varphi_i}\right). \tag{15}
$$

---

[6] As with a regular probit model, γ can only be identified up to a scalar multiple. Let k be a scalar and $k > 0$, according to equations (12) and (13),

$$
\begin{aligned}
& z_i'\gamma + u_i \le 0 \rightarrow kz_i'\gamma + ku_i \le 0 \\
& \rightarrow k\mu_i'\gamma + (ku_i + kv_i'\gamma) \le 0, \text{ and } (ku_i + kv_i'\gamma) \sim N(0,\ k^2(1 + \gamma'\Sigma_i^*\gamma)) \\
& \rightarrow \mathrm{prob}(ku_i + kv_i'\gamma \le -k\mu_i'\gamma) = \Phi\left(\frac{-k\mu_i'\gamma}{k\sqrt{1 + \gamma'\Sigma_i^*\gamma}}\right) \\
& = \Phi\left(\frac{-\mu_i'\gamma}{\varphi_i}\right).
\end{aligned}
$$

Comparing the likelihood function for the probit model with measurement errors to the likelihood function for a regular probit model,[7] the difference between these two likelihood functions is found in the second term, representing the likelihood for non-respondents. When average characteristics from non-respondents' neighborhoods, μ_i, replace the true value of non-respondents' characteristics, z_i, variance is increased from 1 to φ_i^2 (= $1 + \gamma'\Sigma_i\gamma$). By combining the individual survey data with the census data, the original truncated sample becomes a censored sample. However, due to measurement errors, members in the new censored sample are independently but not identically distributed. For respondents, $u_i \sim$ i.i.d. N(0, 1), but for non-respondents, $w_i \sim N(0, \varphi_i^2)$.

3.2.2 Parameter identification in the probit model with measurement errors

It is well known that a measurement errors model suffers from problems of parameters identification (Fuller, 1987). In practice, the probit model with measurement errors derived above suffers the same problems, namely (γ, Σ_i^*) cannot be identified simultaneously.[8] To apply the probit model with measurement errors without further complicating the model, one alternative is to replace Σ_i^* by its consistent estimates.

---

[7] The likelihood function for a regular probit model is

$$L = \prod_{I_i=1} \Phi(z_i'\gamma) \cdot \prod_{I_i=0} \Phi(-z_i'\gamma).$$

[8] Since the number of parameters (elements in Σ_i^*) increases with the sample size, there is an incidental parameters problem.

From census data, there are two candidates that can be chosen to replace Σ_i^*. The first candidate, Σ_i, is a variance-covariance matrix estimated from a sample drawn from the census block for non-respondent i. Typically, $\Sigma_i \neq \Sigma_j$ unless non-respondents i and j live in the same census block.

In contrast to Σ_i whose values vary across non-respondents, the second candidate, Σ, is a constant variance-covariance matrix estimated from a sample drawn from the population. This same constant variance-covariance matrix, Σ, is applied to all the non-respondents.

In practice, Σ and Σ_i can be calculated using the "Public-Use Microdata Samples."[9] Researchers can purchase a 5-percent "Public-Use Microdata Samples," and use this sample to calculate Σ; or the 5-percent sample can be broken down into census blocks[10] and Σ_i can be calculated from each census block.

In terms of empirical results, since both Σ and Σ_i lead to consistent parameter estimates, it is difficult to determine whether Σ or Σ_i will do better. Consequences of using Σ and Σ_i will be examined by Monte Carlo experiments in the next chapter.[11]

---

[9] The "Public-Use Microdata Sample" can be purchased from the U. S. Department of Commerce, Bureau of Census, ph: (301) 763-2005.

[10] An alternative is to purchase a 5-percent "Public-Use Microdata Sample" for each census block.

[11] To simplify notation in the following sections of this chapter, Σ_i is used to represent either Σ or Σ_i.

3.3 A self-selection model with measurement errors and a linear demand equation

A self-selection model with measurement errors is derived in this section which replaces the self-selection equation (a probit model) in a self-selection model by the probit model with measurement errors.

Recall that the self-selection equation with measurement errors is defined as:

(1) For a respondent, $I_i = 1$,

$$z_i'\gamma + u_i > 0,\ u_i \sim \text{i.i.d. } N(0,\ 1). \tag{16}$$

(2) For a non-respondent, $I_i = 0$,

$$\begin{aligned} & z_i'\gamma + u_i \leq 0,\ u_i \sim \text{i.i.d. } N(0,\ 1) \\ & \rightarrow \mu_i'\gamma + w_i \leq 0,\ w_i = u_i + v_i'\gamma, \\ & w_i \sim N(0,\ \varphi_i^2), \text{ and} \\ & \varphi_i^2 = 1 + \gamma'\Sigma_i\gamma. \end{aligned} \tag{17}$$

To derive the self-selection model with measurement errors, assume that individual i's demand for a good (Q) is

$$q_i = x_i'\beta + e_i,\ e_i \sim \text{i.i.d. } N(0,\ \sigma^2), \tag{18}$$

where x_i and β are both mx1 vectors. Further, assume that (e_i, u_i) are distributed jointly as a bivariate normal distribution with a density function

$$g(0,\ 0,\ \Omega), \tag{19}$$

where

$$\Omega = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix} \tag{20}$$

is the variance-covariance matrix, and ρ is the correlation coefficient.

For a respondent, the likelihood is

$$\int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du. \tag{21}$$

For non-respondents, assume that the demand is uncorrelated with the measurement errors (i.e. $Cov(e_i, v_i) = 0$), then (e_i, w_i) are distributed jointly as a bivariate normal distribution with a density function

$$g(0,\ 0,\ \Gamma_i), \tag{22}$$

where

$$\Gamma_i = \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & \varphi_i^2 \end{pmatrix} \tag{23}$$

is the variance-covariance matrix.[12] The likelihood for a non-respondent can then be written as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e,\ w,\ \Gamma_i)\ dw\ de. \tag{24}$$

12
$$\begin{aligned} E(w_i e_i) &= E\left[(u_i + v_i'\gamma)e_i\right] = E\left[[u_i + (z_i - \mu_i)'\gamma]e_i\right] \\ &= E\,(u_i e_i + z_i'\gamma e_i - \mu_i'\gamma e_i) = E\,(u_i e_i) \\ &= \rho\sigma. \end{aligned}$$

Based on equations (16) through (24), the likelihood function for the self-selection model with measurement errors and a linear demand equation is

$$L_L = \prod_{I_i=1} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du$$

$$\prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e,\ w,\ \Gamma_i)\ dw\ de. \tag{25}$$

Consistent estimates for $(\gamma,\ \rho,\ \beta,\ \sigma^2)$ can be obtained by maximizing $\ln(L_L)$.

Comparing the likelihood function for the self-selection model with measurement errors to the likelihood function for the self-selection model with a censored sample (Chapter 2, Section 2.3.2), the difference between these two likelihood functions is found in the second term, representing the likelihood for non-respondents. When average characteristics from non-respondents' neighborhoods, μ_i, replace the true value of non-respondents' characteristics, z_i, variance in the self-selection equation is changed from 1 to φ_i^2 $(= 1 + \gamma'\Sigma_i\gamma)$. In the self-selection model with measurement errors, members in the sample are independently, but no longer identically, distributed.

In addition, compared to the ML estimates from a self-selection model with a truncated sample (Chapter 2, Section 2.3.3), the ML estimates from a self-selection model with measurement errors is more efficient due to the newly introduced information μ_i (the average characteristics from non-respondents' neighborhoods) and Σ_i (the corresponding variance-covariance matrix),

3.4 Generalization for closed-ended questionnaires

The above discussion focuses on the case where the dependent variable in the demand function (q_i) is continuous. However, in many CV studies, the demand responses are not continuous. For example, in many open-ended questionnaires the demand responses are censored (e.g. a Tobit model). On the other hand, surveys using referendum-type (closed-ended) questionnaires produce dichotomized responses. In the following discussion, the demand equation in the self-selection model with measurement errors is modified to allow for qualitative and limited dependent variables. The following models present the case where the demand equation is either a Tobit or a probit-related model.

3.4.1 A self-selection model with measurement errors and a Tobit demand equation

A Tobit demand equation is defined as:

$$\begin{aligned} q_i^* &= x_i'\beta + e_i,\ e_i \sim \text{i.i.d. } N(0,\ \sigma^2), \\ q_i &= q_i^*,\ \text{if } x_i'\beta + e_i > 0, \\ q_i &= 0,\ \text{otherwise.} \end{aligned} \tag{26}$$

The observed demand is now q_i which is left censored at 0. A self-selection model with measurement errors and a Tobit demand equation is described by equations (16), (17), (26), (19), (20), (22), and (23).

For a respondent, if the observed demand equals 0, the likelihood is

$$\int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e,\ u,\ \Omega)\ du\ de. \tag{27}$$

If the observed demand for a respondent is $q_i > 0$, the likelihood is

$$\int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du. \tag{28}$$

For a non-respondent, the likelihood is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e,\ w,\ \Gamma_i)\ dw\ de. \tag{29}$$

Based on equations (27), (28), and (29), the likelihood function for the self-selection model with measurement errors and a Tobit demand equation is

$$\begin{aligned} L_T = & \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e,\ w,\ \Gamma_i)\ dw\ de \\ & \prod_{I_i=1,q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e,\ u,\ \Omega)\ du\ de \\ & \prod_{I_i=1,q_i>0} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du. \end{aligned} \tag{30}$$

In the likelihood function, L_T, the first term is the likelihood for non-respondents. The second term is the likelihood function for those respondents whose $q_i = 0$. The third term is the likelihood function for those respondents whose $q_i > 0$.

3.4.2 A self-selection model with measurement errors and a probit demand equation

In a referendum-type (closed-ended) questionnaire, a respondent is usually asked to answer YES or NO with respect to a given referendum index.[13] The demand equation takes the form

---

[13] For example, a respondent may face a question such as "To maintain the current water quality in your neighborhood, you will have to pay extra $100 per year. Are you willing to pay for it or not?" The $100 here is the referendum index (price).

$$q_i^* = x_i'\beta + e_i,\ e_i \sim \text{i.i.d. } N(0,\ 1),$$
$$q_i = 1, \text{ if } x_i'\beta + e_i > 0, \tag{31}$$
$$q_i = 0, \text{ otherwise},$$

where one of the x_i elements is the referendum index. For respondents, the probit demand equation is related to the self-selection equation (equation (16)) by the assumption that (e_i, u_i) are distributed jointly as a bivariate normal distribution with a density function

$$g(0,\ 0,\ \Theta), \tag{32}$$

where

$$\Theta = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{33}$$

is the variance-covariance matrix.

For a respondent who answers YES with respect to the referendum index, $I_i = 1$ and $q_i = 1$, the likelihood is

$$\int_{-x_i'\beta}^{\infty} \int_{-z_i'\gamma}^{\infty} g(e,\ u,\ \Theta)\ du\ de. \tag{34}$$

For a respondent who answers NO with respect to the referendum index, $I_i = 1$ and $q_i = 0$, the likelihood is

$$\int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e,\ u,\ \Theta)\ du\ de. \tag{35}$$

For non-respondents, the relationship between the probit demand equation and the self-selection equation with measurement errors (equation (17)) can be derived where (e_i, w_i) are distributed jointly as a bivariate normal distribution with density function

$$g(0, 0, \Lambda_i), \tag{36}$$

where

$$\Lambda_i = \begin{pmatrix} 1 & \rho \\ \rho & \varphi_i^2 \end{pmatrix}, \tag{37}$$

and $\varphi_i^2 = 1 + \gamma'\Sigma_i\gamma$. The likelihood for a non-respondent is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e, w, \Lambda_i)\, dw\, de. \tag{38}$$

Based on equations (34), (35), and (38), the likelihood function for the self-selection model with measurement errors and a probit demand equation is

$$\begin{aligned} L_P = &\prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e, w, \Lambda_i)\, dw\, de \\ &\prod_{I_i=1, q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de \\ &\prod_{I_i=1, q_i=1} \int_{-x_i'\beta}^{\infty} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de. \end{aligned} \tag{39}$$

In the likelihood function, L_P, the first term is the likelihood for non-respondents. The second term is the likelihood function for those respondents who answered NO with respect to the referendum index. The third term is the likelihood function for those respondents who answered YES with respect to the referendum index.

An alternative to a probit demand equation is a censored probit inverse demand equation (Cameron and James, 1987; Cameron, 1988).[14] Instead of modeling the probability of answering YES or NO with respect to a referendum index, a censored probit model treats the answer YES (NO) as if q_i^* is greater than or equal to (less than) the referendum index. Thus, the true q_i is censored at the referendum index. However, in terms of econometric estimation, a censored probit demand equation produces results comparable to that of a probit demand equation (McConnell, 1990).

For a self-selection model with measurement errors and a censored probit inverse demand equation, the censored probit inverse demand equation is defined as:

$$
\begin{aligned}
& q_i^* = x_i'\beta + e_i, \; e_i \sim \text{i.i.d. } N(0, \sigma^2), \\
& q_i = 1, \text{ if } x_i'\beta + e_i > p_i, \qquad\qquad (40) \\
& q_i = 0, \text{ otherwise,}
\end{aligned}
$$

where p_i is the referendum index and x_i no longer contains the referendum index. The self-selection model with measurement errors and a censored probit inverse demand equation is defined by equations (16), (17), (40), (19), (20), (22), and (23). It can be easily shown that the likelihood function for the self-selection model with measurement errors and a censored inverse probit demand equation is[15]

---

[14] If the demand equation is estimated by a probit model, the censored probit model estimates the inverse demand equation.

[15] Unlike the case of probit demand equation, in a censored probit (logit) inverse demand equation, σ^2 is identifiable.

$$L_{CP} = \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i' \gamma} g(e,\ u,\ \Gamma_i)\ du\ de$$

$$\prod_{I_i=1,q_i=0} \int_{-\infty}^{p_i - x_i' \beta} \int_{-z_i' \gamma}^{\infty} g(e,\ u,\ \Omega)\ du\ de \tag{41}$$

$$\prod_{I_i=1,q_i=1} \int_{p_i - x_i' \beta}^{\infty} \int_{-z_i' \gamma}^{\infty} g(e,\ u,\ \Omega)\ du\ de.$$

In the likelihood function, L_{CP}, the first term is the likelihood for non-respondents. The second term is the likelihood function for those respondents who answered NO with respect to the referendum index, p_i. The third term is the likelihood function for those respondents who answered YES with respect to the referendum index, p_i.[16]

3.5 Summary

Models derived in this chapter take the average characteristics from the non-respondents' neighborhoods and treat them as the non-respondents' characteristics, measured with error. Based on the measurement errors approach, the probit self-selection equation is modified and becomes a probit model with measurement errors. A self-selection model with measurement errors is constructed using the probit model with measurement errors and a linear demand equation.

CV studies use either open-ended or closed-ended questionnaires to collect data. For open-ended questionnaires, responses to demand are sometimes censored. For example, given a specific price, demand for a good may be left

---

[16] A double-bounded censored logistic regression developed by Hoehn and Loomis (1993) can also be applied. Derivation of the likelihood function is straightforward.

censored at zero. For closed-ended questionnaires, the responses are dichotomized (YES or NO). To account for these situations, the self-selection model with measurement errors is generalized to allow for a Tobit demand equation, a probit demand equation, or a censored probit inverse demand equation.

Based on the measurement errors approach, models derived in this chapter transfer a truncated sample into a censored sample. By applying these models, it is expected that disadvantages from estimates under a truncated sample are removed and advantages from the properties of the estimates under a censored sample are obtained; namely, reliable estimates of the parameters in both the self-selection and the demand equations. Furthermore, some gain in efficiency is expected.

CHAPTER 4

MONTE CARLO EXPERIMENTS AND RESULTS

In the previous chapter, self-selection models with measurement errors were developed with 1) a linear demand equation; 2) a Tobit demand equation; 3) a probit demand equation; and 4) a censored probit inverse demand equation.

Deviating from conventional measurement errors models, the variance-covariance matrix of the measurement errors was replaced by its consistent estimates. Two candidates were considered as replacements for the variance-covariance matrix. One candidate, Σ, was the variance-covariance matrix estimated from a sample drawn from the population, and was not available for each census block. The other candidate, Σ_i, was the variance-covariance matrix estimated from samples drawn from each non-respondent's census block.[1]

The purpose of this chapter is to use Monte Carlo experiments to examine and compare the resulting estimates from 1) a truncated sample without correcting for self-selection bias; 2) a self-selection model with a censored sample;[2] 3) a self-selection model with measurement errors that adopts μ_i, the

[1] A third type of variance-covariance diag(Σ_i) which assumes zero covariance was also tried. Although the diag(Σ_i) is very easy to obtain, it is abandoned for two reasons. First, the zero covariance assumption is not plausible. Second, according to the model specified below, ML estimator based on diag(Σ_i) has never converged during the optimization procedure.

[2] Although it is nearly impossible to acquire a censored sample in reality, estimates from a censored sample give the best possible results and can be used to compare the results from the measurement errors models proposed in this study.

mean vector, and Σ; and 4) a self-selection model with measurement errors that adopts μ_i and Σ_i.[3] Monte Carlo experiments are conducted for each type of demand equation except the censored probit inverse demand equation.[4]

This chapter begins with the data generation process. Steps for Monte Carlo experiments are described and the resulting estimates are then reported. Comparison of the results are presented, followed by concluding remarks.

4.1 Data generation

Due to the properties of the proposed self-selection models with measurement errors, the data generation process is not straightforward. In each replication, in order to acquire useful information, data used in Monte Carlo experiments are generated in two steps. In the first step, a "population" is generated and certain required statistics are calculated. In the second step, a "sample" is drawn from the "population," and models are estimated based on the "sample."

---

[3] Monte Carlo experiments for a self-selection model with a truncated sample is conducted only for a linear demand equation with $\rho = 0.75$.

[4] Due to the similarity between a probit and a censored probit model, estimates from a censored probit model are omitted.

4.1.1 Population generation

For each replication, a 10,000 x 5 matrix, $[x_1\ x_2\ x_3\ u\ e]$, is first generated where $[x_{i1}\ x_{i2}\ x_{i3}\ u_i\ e_i]$ is distributed as an i.i.d. multivariate normal distribution with a mean vector [3 1.5 4 0 0] and a variance-covariance matrix[5]

$$\mathrm{Cov}(x_{i1}, x_{i2}, x_{i3}, u_i, e_i) = \begin{bmatrix} 1.44 & 0.24 & 0.096 & 0 & 0 \\ 0.24 & 1 & 0.24 & 0 & 0 \\ 0.096 & 0.24 & 0.64 & 0 & 0 \\ 0 & 0 & 0 & 1 & \rho \\ 0 & 0 & 0 & \rho & 1 \end{bmatrix},$$

where ρ = 0.25, 0.5, or 0.75.[6]

Since one of the demand specifications and the self-selection equation are both probit equations, setting Var(u) = Var(e) = 1 simplifies comparison of parameter estimates.[7]

Dependent variables for both the self-selection equation (I_i^*) and the demand equation (q_i^*) are generated by

---

[5] $\mathrm{Corr}(x_1, x_2) = 0.2$, $\mathrm{Corr}(x_1, x_3) = 0.1$, $\mathrm{Corr}(x_2, x_3) = 0.3$, $\mathrm{Corr}(u, e) = \rho$, and ρ = 0.25, 0.5 or 0.75.

[6] Based on ρ = 0.25, 0.5, and 0.75, three sequences of simulations are conducted for each of the models.

[7] Recall that in a probit model, β and σ are not separately identifiable. Coefficients estimated are β/σ.

$$I_i^* = 1.5 + 1\, x_{i1} - 3\, x_{i2} + u_i, \text{ and}$$

$$q_i^* = 6 + 4\, x_{i2} - 3\, x_{i3} + e_i.$$

In order for the model to be identifiable when both demand and self-selection equations are of probit-type, both demand and self-selection equations cannot have exactly the same independent variables.[8]

Based on the process described above, a sample [I^* q^* x_1 x_2 x_3 u e], which contains 10,000 observations and 7 variables, is generated and treated as the "population" in a replication.

To apply the self-selection models with measurement errors, certain statistics related to the distribution of x_1 and x_2 are required (i.e. the mean vector and variance-covariance matrix). To obtain the necessary statistics, a random sample containing 200 observations is drawn from the "population" and the variance-covariance matrix (Σ) of x_1 and x_2 is calculated. The next step is to randomly group the "population" into 250 "blocks" with 40 observations in each block. For each block, the mean vector (μ_i) and the variance-covariance matrix (Σ_i) of x_1 and x_2 are calculated.

4.1.2 Sample generation

In each replication, a random sample consisting of 1,000 observations is drawn from the population. In the random sample, observations with $I_i^* > 0$ ($I_i^* \leq 0$) are treated as respondents (non-respondents). Since the mean of I_i^* is zero, a response rate roughly equaling 50% (500 respondents) is expected.

---

[8] An alternative is to have Corr(u, e) = 0. However, if this is the case, self-selection does not exist.

In each replication, four models are estimated: 1) without correcting for self-selection bias, a demand equation is estimated based on the truncated sample, i.e. the number of observations is about 500; 2) both a demand and a self-selection equation are estimated based on a censored sample with 1,000 observations; i.e. for non-respondents, x_1 and x_2 are observable; 3) both a demand and a self-selection equation are estimated using a self-selection model with measurement errors, and for non-respondents, due to the unobserved x_1 and x_2, μ_i and Σ are used (i.e. the number of observations is 1,000); and 4) both a demand and a self-selection equation are estimated using a self-selection model with measurement errors, and for non-respondents, due to the unobserved x_1 and x_2, μ_i and Σ_i are used (i.e. the number of observations is 1,000).

As previously mentioned, three types of demand equations are used in the analysis. For a linear demand equation, q_i^* is used as the dependent variable. If the demand equation is a Tobit equation, q_i^* is left censored at 0 ($q_i = q_i^*$, if $q_i^* > 0$; 0, otherwise). Finally, for a probit demand equation, q_i^* is dichotomized ($q_i = 1$, if $q_i^* > 0$; 0, otherwise).

4.1.3 Monte Carlo experiments

Based on different demand specifications, three types of simulations related to a linear, a Tobit, and a probit demand equation are conducted. For each type of demand specification, three sequences of simulations are conducted based on different values of the correlation between self-selection and demand ($\rho = 0.25$, 0.5, and 0.75). At each replication, four models are estimated, and the number of replications is 500.

4.2 A linear demand equation with self-selection

A self-selection model with measurement errors and a linear demand equation is derived in Chapter 3 (Section 3.3). Based on the different correlation measures between self-selection and demand ($\rho = 0.25$, 0.5, and 0.75), the following section begins with OLS estimates from a truncated sample without correcting for self-selection, and results are presented in Appendix C[9] (Tables C.1.1.A ($\rho = 0.25$), C.1.2.A ($\rho = 0.5$), and C.1.3.A ($\rho = 0.75$)).

Using a censored sample, estimates for a linear demand equation with self-selection are obtained by the ML estimator based on the likelihood function

$$L_{L1} = \prod_{I_i=1} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du$$

$$\prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-z_i'\gamma} g(e,\ u,\ \Omega)\ du\ de.$$

where g(.,.,.) represents a bivariate normal density function and

$$\Omega = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

is the variance-covariance matrix. Results of this model are listed in Tables C.1.1.B ($\rho = 0.25$), C.1.2.B ($\rho = 0.5$), and C.1.3.B ($\rho = 0.75$).

For a truncated sample and $\rho = 0.75$, estimates for a linear demand equation with self-selection are obtained by the ML estimator based on the likelihood function

---

[9] Notation used in Appendix C are defined in Appendix B.

$$L_t = \prod_{I_i=1} \frac{\int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\, u,\, \Omega)\, du}{\Phi(z_i'\gamma)},$$

and results are listed in Table C.1.3.E.

Estimates from a self-selection model with measurement errors and a linear demand equation are obtained by the ML estimator based on (μ_i, Σ) and the likelihood function

$$L_{L2} = \prod_{I_i=1} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\, u,\, \Omega)\, du$$

$$\prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e,\, w,\, \Gamma)\, dw\, de,$$

where

$$\Gamma = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & (1 + \gamma'\Sigma\gamma) \end{bmatrix},$$

and results are presented in Tables C.1.1.C ($\rho = 0.25$), C.1.2.C ($\rho = 0.5$), and C.1.3.C ($\rho = 0.75$).

Finally, if Σ (Γ) is replaced by Σ_i (Γ_i), i.e.

$$\Gamma_i = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & (1 + \gamma'\Sigma_i\gamma) \end{bmatrix},$$

the resulting estimates are shown in Tables C.1.1.D ($\rho = 0.25$), C.1.2.D ($\rho = 0.5$), and C.1.3.D ($\rho = 0.75$).

4.2.1 Monte Carlo experiment results from a self-selection model with measurement errors and a linear demand equation[10]

Based on Tables C.1.1.A, C.1.2.A, and C.1.3.A, when the demand equation is estimated by applying OLS to a single equation without correcting for self-selection bias, as ρ increases, both %BIAS and $D_{(\beta,\sigma^2);S}$ increase. This implies that the higher the ρ, the farther the OLS results deviate from the true parameter values. For example, as ρ increased from 0.25 to 0.75, the %BIAS of β_1 (σ^2) increased from 2.19% (1.32%) to 6.45% (7.73%). In addition, RMSE and ASE are very different for β_1, indicating incorrect estimates of the variance-covariance matrix.

When a censored sample is available and the self-selection model is correctly specified, σ^2, ρ, self-selection, and demand parameters are well-estimated by the ML estimator. As can be seen from Tables C.1.1.B, C.1.2.B, and C.1.3.B, the %BIAS among demand (self-selection) parameters ranged from 0.01% (0.08%) to 0.28% (1.20%); for σ^2 (ρ), %BIAS ranged from 0.32% (0.28%) to 0.62% (1.08%). $D_{(\beta,\sigma^2);CEN}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);CEN}$'s remained very close to zero.

For a truncated sample, the ML estimator produces different results from that of Muthén and Jöreskog (1983). Table C.1.3.E shows that biasedness is not a major problem for all the σ^2 ,ρ, self-selection, and demand parameters, even with ρ = 0.75. The real problem appears to be the difference between RMSE and ASE. The difference between RMSE and ASE indicates that the variance-covariance matrix produced by the ML estimator is incorrect and cannot be used

---

[10] A GAUSS program for conducting the Monte Carlo experiments is provided in Appendix D.

to test hypotheses. Failure to conduct hypothesis testing may result in model misspecification and lead to inconsistent parameter estimates.

When the measurement errors model based on μ_i and Σ was applied, the %BIAS among demand (self-selection) parameters ranged from 0.02% (1.53%) to 0.29% (3.38%); for σ^2 (ρ), %BIAS ranged from 0.28% (0.53%) to 0.80% (0.88%) as shown in Tables C.1.1.C, C.1.2.C, and C.1.3.C. $D_{(\beta,\sigma^2);ME1}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);ME1}$'s remained very close to zero.

When the measurement errors model based on μ_i and Σ_i was applied, the %BIAS among demand (self-selection) parameters ranged from 0.01% (0.15%) to 0.28% (2.43%); for σ^2 (ρ), %BIAS ranged from 0.22% (0.40%) to 0.68% (0.82%) as shown in Tables C.1.1.D, C.1.2.D, and C.1.3.D. $D_{(\beta,\sigma^2);ME2}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);ME2}$'s remained very close to zero.

Comparing results from the two measurement errors models, the only difference is that the self-selection parameters always have smaller %BIAS when Σ_i is used. Apart from this, it is difficult to distinguish the difference between the two models.

Comparing results from the two measurement errors models with results from the censored sample, all three models give similar estimates for the demand parameters according to $D_{(\beta,\sigma^2);\cdot}$. However, according to $D_{(\gamma,\rho);\cdot}$, self-selection parameters estimated by the two measurement errors models are less efficient than the estimates from the censored sample.

4.3 A Tobit demand equation with self-selection

A self-selection model with measurement errors and a Tobit demand equation is derived in Chapter 3 (Section 3.4.1). Based on the different correlation measures between self-selection and demand (ρ = 0.25, 0.5, and 0.75), the following section begins with Tobit ML estimates from a truncated sample without correcting for self-selection, and results are presented in Appendix C (Tables C.2.1.A (ρ = 0.25), C.2.2.A (ρ = 0.5), and C.2.3.A (ρ = 0.75)).

Using a censored sample, estimates for a Tobit demand equation with self-selection are obtained by the ML estimator based on the likelihood function

$$L_{T1} = \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-z_i'\gamma} g(e,\ u,\ \Omega)\ du\ de$$

$$\prod_{I_i=1, q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e,\ u,\ \Omega)\ du\ de$$

$$\prod_{I_i=1, q_i>0} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta,\ u,\ \Omega)\ du,$$

where g(.,.,.) is a bivariate normal density function and

$$\Omega = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}$$

is the variance-covariance matrix. Results of this model are listed in Tables C.2.1.B (ρ = 0.25), C.2.2.B (ρ = 0.5), and C.2.3.B (ρ = 0.75).[11]

Estimates from a self-selection model with measurement errors and a Tobit demand equation are obtained by the ML estimator based on (μ_i, Σ) and the likelihood function

---

[11] A Tobit self-selection model based on a truncated sample is dropped from the Monte Carlo experiments due to the difficulty in obtaining the starting values. The ML estimator for a Tobit self-selection model based on a truncated sample is very sensitive to the starting values. Very often, the optimization procedure can not converge even with the true parameter values as the starting values.

$$L_{T2} = \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e, w, \Gamma)\, dw\, de$$

$$\prod_{I_i=1, q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e, u, \Omega)\, du\, de$$

$$\prod_{I_i=1, q_i>0} \int_{-z_i'\gamma}^{\infty} g(q_i - x_i'\beta, u, \Omega)\, du,$$

where

$$\Gamma = \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & (1 + \gamma'\Sigma\gamma) \end{bmatrix},$$

and results are presented in Tables C.2.1.C ($\rho = 0.25$), C.2.2.C ($\rho = 0.5$), and C.2.3.C ($\rho = 0.75$).

Finally, the results of replacing Σ (Γ) by Σ_i (Γ_i) are shown in Tables C.2.1.D ($\rho = 0.25$), C.2.2.D ($\rho = 0.5$), and C.2.3.D ($\rho = 0.75$).

4.3.1 Monte Carlo experiment results from a self-selection model with measurement errors and a Tobit demand equation[12]

When a single equation Tobit model is applied to estimate the demand equation without correcting for self-selection bias, as in the case of a linear demand equation, both %BIAS and $D_{(\beta,\sigma^2);S}$ increase with ρ as shown in Tables C.2.1.A, C.2.2.A, and C.2.3.A. This again implies that the higher the ρ, the farther the estimates from a single equation Tobit model deviate from the true parameter values. For example, as ρ increased from 0.25 to 0.75, the %BIAS of

---

[12] A GAUSS program for conducting the Monte Carlo experiments is provided in Appendix E.

β_1 (σ^2) increased from 3.84% (2.26%) to 11.09% (12.33%). In addition, RMSE and ASE are very different for β_1, indicating incorrect estimates of the variance-covariance matrix.

When a censored sample is available and the self-selection model is correctly specified, σ^2, ρ, self-selection, and demand parameters are well-estimated by the ML estimator. As can be seen from Tables C.2.1.B, C.2.2.B, and C.2.3.B, the %BIAS among demand (self-selection) parameters ranged from 0.09% (0.07%) to 0.30% (2.19%); for σ^2 (ρ), %BIAS ranged from 0.12% (1.39%) to 1.46% (3.28%). $D_{(\beta,\sigma^2);CEN}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);CEN}$'s remained very close to zero.

When measurement errors model based on μ_i and Σ was applied, the %BIAS among demand (self-selection) parameters ranged from 0.08% (1.75%) to 0.35% (4.68%); for σ^2 (ρ), %BIAS ranged from 0.18% (1.32%) to 1.42% (2.92%) as shown in Tables C.2.1.C, C.2.2.C, and C.2.3.C. $D_{(\beta,\sigma^2);ME1}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);ME1}$'s remained very close to zero.

When the measurement errors model based on μ_i and Σ_i was applied, the %BIAS among demand (self-selection) parameters ranged from 0.07% (1.23%) to 0.36% (3.71%); for σ^2 (ρ), %BIAS ranged from 0.16% (1.37%) to 1.38% (3.06%) as can be seen in Tables C.2.1.D, C.2.2.D, and C.2.3.D. $D_{(\beta,\sigma^2);ME2}$ was always smaller than $D_{(\beta,\sigma^2);S}$ and all the $D_{(\cdot);ME2}$'s remained very close to zero.

Comparing results from the two measurement errors models, the only difference is that the self-selection parameters always have smaller %BIAS when Σ_i is used. Apart from this, it is difficult to distinguish the difference between the two models.

Comparing results from the two measurement errors models with results from the censored sample, all three models give similar estimates for the demand parameters according to $D_{(\beta,\sigma^2);\cdot}$. However, according to $D_{(\gamma,\rho);\cdot}$, self-selection parameters estimated by the two measurement errors models are less efficient than the estimates from the censored sample.

4.4 A probit demand equation with self-selection

A self-selection model with measurement errors and a probit demand equation is derived in Chapter 3 (Section 3.4.2). Based on the different correlation measures between self-selection and demand ($\rho = 0.25$, 0.5, and 0.75), the following section begins with probit ML estimates from a truncated sample without correcting for self-selection, and results are presented in Appendix C (Tables C.3.1.A ($\rho = 0.25$), C.3.2.A ($\rho = 0.5$), and C.3.3.A ($\rho = 0.75$)).

Using a censored sample, estimates for a probit demand equation with self-selection are obtained by the ML estimator based on the likelihood function

$$L_{P1} = \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-z_i'\gamma} g(e, u, \Theta)\, du\, de$$

$$\prod_{I_i=1, q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de$$

$$\prod_{I_i=1, q_i=1} \int_{-x_i'\beta}^{\infty} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de,$$

where

$$\Theta = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

is the variance-covariance matrix. Results of this model are listed in Tables C.3.1.B ($\rho = 0.25$), C.3.2.B ($\rho = 0.5$), and C.3.3.B ($\rho = 0.75$).[13]

Estimates from a self-selection model with measurement errors and a probit demand equation are obtained by the ML estimator based on (μ_i, Σ) and the likelihood function

$$L_{P2} = \prod_{I_i=0} \int_{-\infty}^{\infty} \int_{-\infty}^{-\mu_i'\gamma} g(e, w, \Lambda)\, dw\, de$$

$$\prod_{I_i=1, q_i=0} \int_{-\infty}^{-x_i'\beta} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de$$

$$\prod_{I_i=1, q_i=1} \int_{-x_i'\beta}^{\infty} \int_{-z_i'\gamma}^{\infty} g(e, u, \Theta)\, du\, de,$$

where

$$\Lambda = \begin{bmatrix} 1 & \rho \\ \rho & (1+\gamma'\Sigma\gamma) \end{bmatrix},$$

and results are presented in Tables C.3.1.C ($\rho = 0.25$), C.3.2.C ($\rho = 0.5$), and C.3.3.C ($\rho = 0.75$).

Finally, if Σ (Λ) is replaced by Σ_i (Λ_i), i.e.

$$\Lambda_i = \begin{bmatrix} 1 & \rho \\ \rho & (1+\gamma'\Sigma_i\gamma) \end{bmatrix},$$

---

[13] A probit self-selection model based on a truncated sample is dropped from the Monte Carlo experiments due to the difficulty in obtaining the starting values. The ML estimator for a probit self-selection model based on a truncated sample is very sensitive to the starting values. Very often, the optimization procedure can not converge even with the true parameter values as the starting values.

the resulting estimates are shown in Tables C.3.1.D (ρ = 0.25), C.3.2.D (ρ = 0.5), and C.3.3.D (ρ = 0.75).

4.4.1 Monte Carlo experiment results from a self-selection model with measurement errors and a probit demand equation[14]

When a single equation probit model is applied to estimate the demand equation without correcting for self-selection bias, as in the case of a linear demand equation, both %BIAS and $D_{(\beta,\sigma^2);S}$ increase with ρ as presented in Tables C.3.1.A, C.3.2.A, and C.3.3.A. This again implies that the higher the ρ, the farther the estimates from a single equation probit model deviate from the true parameter values. For example, as ρ increased from 0.25 to 0.75, the %BIAS of β_1 increase from 7.43% to 18.35%. In addition, RMSE and ASE are very different for β_1, indicating incorrect estimates of the variance-covariance matrix.

When a censored sample is available and the self-selection model is correctly specified, ρ, self-selection, and demand parameters are well-estimated by the ML estimator. As can be seen from Tables C.3.1.B, C.3.2.B, and C.3.3.B, the %BIAS among demand (self-selection) parameters ranged from 2.63% (0.45%) to 4.63% (2.19%); for σ^2 (ρ), %BIAS ranged from 2.97% (1.08%) to 3.60% (10.28%). $D_{\beta;CEN}$ was smaller than $D_{\beta;S}$ when ρ = 0.5 and 0.75 and all the $D_{(\cdot);ME1}$'s remained very close to zero.

When measurement errors model based on μ_i and Σ was applied, the %BIAS among demand (self-selection) parameters ranged form 0.04% (3.46%) to 4.84% (7.43%); for σ^2 (ρ), %BIAS ranged from 3.04% (1.36%) to 3.74% (9.71%)

[14] A GAUSS program for conducting the Monte Carlo experiments is provided in Appendix F.

as shown in Tables C.3.1.C, C.3.2.C, and C.3.3.C. $D_{\beta;ME1}$ was smaller than $D_{\beta;S}$ when $\rho = 0.5$ and 0.75 and all the $D_{(\cdot);CEN}$'s remained very close to zero.

When the measurement errors model based on μ_i and Σ_i was applied, the %BIAS among demand (self-selection) parameters ranged from 2.56% (2.26%) to 4.69% (4.97%); for σ^2 (ρ), %BIAS ranged from 3.00% (1.16%) to 3.65% (9.99%) as presented in Tables C.3.1.D, C.3.2.D, and C.3.3.D. $D_{\beta;ME2}$ was smaller than $D_{\beta;S}$ when $\rho = 0.5$ and 0.75 and all the $D_{(\cdot);ME2}$'s remained very close to zero.

Comparing results from the two measurement errors models, the only difference is that the self-selection parameters always have smaller %BIAS when Σ_i is used. Apart from this, it is difficult to distinguish the difference between the two models.

Comparing results from the two measurement errors models with results from the censored sample, all three models give similar estimates for the demand parameters according to $D_{\beta;\cdot}$. However, according to $D_{(\gamma,\rho);\cdot}$, self-selection parameters estimated by the two measurement errors models are less efficient than the estimates from the censored sample.

One important issue is the estimate of ρ. For all three self-selection models, as the true value of ρ increases, the %BIAS for the estimate increases rapidly. However, BIAS for the estimates of ρ are always equal to zero, statistically.

4.5 General results from the Monte Carlo experiments

Results from the single equation simulation show that in the presence of self-selection ($\rho \neq 0$), %BIAS increases as ρ increases when a single equation is used to estimate the demand equation. This indicates biasedness caused by the self-selection behavior.

For a truncated sample, the ML estimator produces different results from that of Muthén and Jöreskog (1983). Instead of biasedness, the real problem appears to be that the variance-covariance matrix produced by the ML estimator is incorrect and cannot be used to test hypotheses. Failure to conduct hypothesis testing may result in model misspecification and may lead to inconsistent parameter estimates.

When a censored sample is available and the self-selection model is correctly specified, σ^2, self-selection, and demand parameters are well-estimated by the ML estimator. For the parameter ρ, the ML estimator leads to acceptable results; however, the estimates are not as accurate as other parameter estimates, especially in the case of a probit demand equation with self-selection.

In terms of efficiency among different estimators, $D_{(\beta,\sigma^2);CEN}$ being very close to that of $D_{(\beta,\sigma^2);ME1}$ and $D_{(\beta,\sigma^2);ME2}$[15] indicates that the model which uses a censored sample and the two measurement errors models all lead to very similar estimates of demand parameters and σ^2. For the self-selection parameters and ρ, $D_{(\gamma,\rho);ME1} > D_{(\gamma,\rho);ME2} > D_{(\gamma,\rho);CEN}$, indicates that the model which uses a censored sample performs the best and the measurement errors model that uses μ_i and Σ_i performs somewhat better than the model that uses μ_i and Σ. In the overall performance, it is no surprise that the model which uses a censored sample has the smallest value of $D_{(\gamma,\beta,\sigma^2,\rho);CEN}$ and performs the best. Even though the value of $D_{(\gamma,\beta,\sigma^2,\rho);ME1}$ is slightly greater than that of $D_{(\gamma,\beta,\sigma^2,\rho);ME2}$, the two measurement errors models are not very different from each other.

There is a problem common to the case of the self-selection model with a probit demand equation. The model that uses censored sample or either

---

[15] For the probit demand equation case, they are $D_{\beta;CEN}$, $D_{\beta;ME1}$ and $D_{\beta;ME2}$ respectively. In the following discussion, (β, σ^2) is used to represent (β) in the probit demand case as well as (β, σ^2) in other cases.

measurement errors model results in an estimate of ρ that is not as accurate as other parameters, especially when the true value of ρ is high. However, the estimate remains statistically acceptable.

4.6 Summary

In this chapter, Monte Carlo experiments are conducted to examine and to compare the resulting estimates from 1) a truncated sample without correcting for self-selection bias; 2) a self-selection model with a censored sample; 3) a self-selection model with measurement errors that adopts μ_i, the mean vector, and Σ, the corresponding variance-covariance matrix estimated from a sample drawn from the population; and 4) a self-selection model with measurement errors that adopts μ_i and Σ_i, the corresponding variance-covariance matrix estimated from samples drawn from each non-respondent's census block. Three sequences of Monte Carlo experiments are conducted based on a linear, a Tobit, and a probit demand equation. For each sequence of Monte Carlo experiment, based on ρ = 0.25, 0.5, and 0.75, three 500-replication simulations are executed.

Results from the Monte Carlo experiments show that the ML estimator from the model which uses a censored sample performs the best. Among the two measurement errors models, the model that uses μ_i and Σ_i estimates the self-selection parameters more accurately than the model that uses μ_i and Σ.

In reality, censored sample is almost impossible to obtain. However, using the measurement errors models derived in this study, a truncated sample can be transferred into a censored sample, and self-selection models with measurement errors can then be estimated by ML estimators. Results from Monte Carlo experiments show that the estimates from the self-selection models with measurement errors perform very well. According to the Monte Carlo experiment results, when a correctly-specified self-selection model with measurement errors is adopted, estimates of the demand parameters are as accurate as the estimates from a model that uses a censored sample, and the estimates of the self-selection parameters are very close to the true parameter values.

The results indicate an impressive message: adoption of a self-selection model with measurement errors will not contaminate the original truncated sample. Compared to the estimates from a model with truncated sample, the self-selection models with measurement errors not only improve the efficiency of the estimates but also lead to reliable estimates of the self-selection parameters.

CHAPTER 5

CONCLUDING REMARKS

In CV studies, when surveys are used for collecting data, non-response will usually create problems. In analyzing survey data, two types of possible biases can be created by non-response. The first is sample non-response bias which occurs when the sample distribution of some socio-economic or demographic characteristics is significantly different from that of the population. The second is self-selection bias which occurs when the non-response is non-random, i.e. the reasons for non-response are endogenous to the survey study.

In CV studies, although self-selection is usually ignored in empirical work, it is recognized by researchers as an important issue. In this study, methods that combine survey individual data with census data to correct for self-selection bias are proposed and promising results are provided by Monte Carlo experiments.

5.1 Summary

In Chapter 1, consequences of self-selection are reported and the differences between self-selection bias and sample non-response bias are distinguished. When regression is used to analyze survey data, it is shown that self-selection causes inconsistent parameter estimates and sample non-response bias does not even play a role. It is also shown that there is no direct relationship between sample non-response bias and self-selection bias. Instead of ignoring self-selection bias in empirical work, it is suggested that CV researchers treat self-selection as a serious issue.

Following an example in labor economics, the concept of self-selection in CV is introduced in Chapter 2. It is identified that a complete self-selection model consists of two equations. The first is a self-selection equation which is essentially a probit equation, and the second is a demand equation. To estimate a self-selection model, several estimators that simultaneously estimate the self-selection equation and the demand equation have been reviewed. However, because the CV survey data is a truncated sample, evidence shows that the self-selection equation parameters cannot be estimated reliably by existing estimators. It is the deficiency of existing estimators that motivates this study.

In Chapter 3, a self-selection model under a CV framework is derived and new ML estimators are proposed. According to a random utility model, a self-selection equation can be expressed by a probit model with income as one of the important explanatory variables. A self-selection model is completely described by a self-selection probit equation and a demand equation which is correlated with the self-selection equation. Since a CV data set is usually a truncated sample where the only information available for a non-respondent is the address, a self-selection model with measurement errors is derived by combining the CV truncated sample with census data which provides information for non-respondents' neighborhoods (e.g. census blocks). Based on the self-selection model with measurement errors, two ML estimators are then proposed. Finally, the self-selection model with measurement errors is extended to allow for a demand equation with qualitative or limited dependent variables.

It is found in Chapter 4 that for a truncated sample, the ML estimator produces different results from that of Muthén and Jöreskog (1983). Biasedness is not a major problem even with $\rho = 0.75$. The real problem appears to be the difference between RMSE and ASE which indicates that the variance-covariance matrix produced by the ML estimator is incorrect and cannot be used in testing hypotheses. If the self-selection equation cannot be correctly specified, all of the σ^2, ρ, self-selection, and demand parameters may be estimated inconsistently.

The main purpose of Chapter 4 is to use Monte Carlo experiments to compare the resulting parameter estimates from the two ML estimators for the self-selection models with measurement errors proposed in Chapter 3. For the first estimator, a sample drawn from the population is used to calculate the variance-covariance matrix (Σ) for non-respondents' explanatory variables in self-selection equation, and for the second estimator, the variance-covariance matrix for non-respondents' explanatory variables is calculated using samples drawn from each non-respondent's census block (Σ_i). Monte Carlo results show that both of the ML estimators give very accurate estimates for all of the self-selection and demand parameters. However, in terms of efficiency, the estimator using Σ_i performs somewhat better than the estimator using Σ.

Although the ML estimator using Σ_i performs only slightly better than the alternative estimator using Σ, it is the estimator that is recommended. Consider a case where census blocks are heterogeneous ($\Sigma_i^* \neq \Sigma_j^*$, $i \neq j$). In this case, Σ is no longer a consistent estimate for Σ_i^*, and the resulting estimator using Σ does not lead to consistent parameter estimates. Although Σ is easier to obtain and performs similarly to Σ_i, a stronger assumption is needed to assure the consistency of parameter estimates.

5.2 Need for future research

It is indicated in Chapter 4 (Section 4.2.1) that Monte Carlo experiment results from the ML estimator based on a truncated sample are different from that of Muthén and Jöreskog (1983). The reasons behind this difference remain to be explored by future studies that concentrate on the issue of model specification. It is important to determine the degree to which self-selection models with measurement errors are sensitive to model misspecification.

Another area that remains to be explored is the large %BIAS and incorrect variance for β_1 that results from the use of a single equation approach without correcting for self-selection bias. This problem may be approached by varying the variance-covariance matrix structure for $[x_{i1}\ x_{i2}\ x_{i3}\ u_i\ e_i]$ and examine how it affects the estimates from a single equation method such as OLS.

In a comparison of the two self-selection models with measurement errors, Monte Carlo experiment results suggest that adoption of μ_i and Σ_i produces better results. Recall that both μ_i and Σ_i are estimated from non-respondent i's neighborhood, and the neighborhood is loosely defined as a census block, a county, a state, or even a region. Definition of the neighborhood remains an empirical problem and should be studied further.

5.3 Conclusion

In CV studies, data can only be collected from those who are willing to participate in the studies. Results from the application of a single equation approach to this truncated sample may lead to inconsistent parameter estimates (self-selection bias). Unfortunately, there is no simple method to detect the existence of self-selection bias in CV studies. A self-selection model which contains a self-selection and a demand equation must be specified in order to detect and to correct for self-selection bias. The ML estimator that is based on the self-selection model with a truncated sample provides theoretically consistent parameter estimates. However, unless the data is a censored sample, it is shown that the parameters and the variance-covariance matrix in the self-selection equation cannot be estimated reliably.

A method that transfers a truncated sample to a censored sample by combining survey individual data and census data is proposed and is called a self-selection model with measurement errors. Two ML estimators are derived based on the self-selection model with measurement errors where data from census are treated as if they are the true values plus errors.

Results from the Monte Carlo experiments show that the ML estimator based on the model which uses a censored sample has the best performance. ML estimators based on the self-selection models with measurement errors perform very well, especially in estimating demand parameters. According to the Monte Carlo experiment results, when a correctly-specified self-selection model with measurement errors is adopted, estimates of the demand parameters are as accurate and efficient as the estimates from a model that uses a censored sample, and the estimates of the self-selection parameters are very close to the true parameter values. The results indicate an impressive message: adoption of a self-selection model with measurement errors will not contaminate the original truncated sample.

Among the two ML estimators from the self-selection model with measurement errors, the model that uses μ_i and Σ_i estimates the self-selection parameters more accurately and efficiently than the model that uses μ_i and Σ. Although Σ is easier to obtain and the ML estimator based on μ_i and Σ produces acceptable results, compared to the estimator that uses μ_i and Σ_i, stronger assumptions are needed to justify the results.

Although self-selection models with measurement errors developed in this study started from a CV study using mail surveys with different demand specifications, they can be easily generalized in several ways. First, since derivation of the model requires no specific restriction for the surveys, any type of cross-section survey can be applied. Second, although a demand function is used in the model, the important issue is the correlation between the demand function and the self-selection equation. The model is still valid even if the demand function is replaced with a supply function, given that it is correctly specified. In general, models developed in this study are broad enough to be applied to studies that adopt survey data and regression analyses.

APPENDIX A

RESULTS FROM MUTHÉN AND JÖRESKOG'S STUDY

APPENDIX A

RESULTS FROM MUTHÉN AND JÖRESKOG'S STUDY

In the model 1 of Muthén and Jöreskog's study (1983, Section 5), the selection relation is specified as:

$$y_i = 0.0 + 1.0\, x_i + \varepsilon_i,$$

$$\eta_i = 0.0 - 1.0\, x_i + \delta_i,$$

where $[x_i\ \varepsilon_i\ \delta_i]$ is distributed as an i.i.d. trivariate normal distribution with a mean vector [0 0 0] and a variance-covariance matrix

$$\mathrm{Cov}(x_i,\ \varepsilon_i,\ \delta_i) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -0.5 \\ 0 & -0.5 & 1 \end{bmatrix}.$$

Based on different sample sizes (i.e. N = 1,000 and N = 4,000), Monte Carlo experiment results are presented in the following tables:[1]

[1] Estimates from a probit and a Heckman's two-stage estimator that were reported by Muthén and Jöreskog are omitted here. Notation used to report results are defined in Appendix B.

Table A.1 Parameter estimates for data simulated according to model 1, N_t = 496,* N = 1000

Parameters	OLS Estimates	ML Estimates	BIAS	%BIAS
		Truncated Sample		
β_0 = 0.0	-.373	-.209	0.209	—
	(.054)**	(.119)		
β_1 = 1.0	.788	.931	0.069	6.9%
	(.052)	(.095)		
$\sigma_{\varepsilon\varepsilon}$ = 1.0	.985	.982	0.018	1.8%
	(.065)	(.076)		
γ_0 = 0.0		.991	0.991	—
		(1.599)		
γ_1 = −1.0		-3.448	2.448	244.8%
		(4.542)		
ρ = −0.5		-.248	0.252	50.4%
		(.413)		
		Censored Sample		
β_0 = 0.0	-.373	.074	0.074	—
	(.054)	(.179)		
β_1 = 1.0	.788	1.033	0.033	3.3%
	(.052)	(.114)		
$\sigma_{\varepsilon\varepsilon}$ = 1.0	.985	1.126	0.126	12.6%
	(.065)	(.131)		
γ_0 = 0.0		.013	0.013	—
		(.046)		
γ_1 = −1.0		-1.040	0.040	4.0%
		(.068)		
ρ = −0.5		-.522	0.022	4.4%
		(.164)		

* Truncated sample size.

** Standard errors in parentheses.

Table A.2 Parameter estimates for data simulated according to model 1, N_t = 1963,* N = 4000

Parameters	OLS Estimates	ML Estimates	BIAS	%BIAS
		Truncated Sample		
β_0 = 0.0	-.435	-.223	0.223	—
	(.027)	(.137)		
β_1 = 1.0	.807	.965	0.035	3.5%
	(.027)	(.084)		
$\sigma_{\varepsilon\varepsilon}$ = 1.0	.916	.978	0.022	2.2%
	(.029)	(.056)		
γ_0 = 0.0		.851	0.851	—
		(.723)		
γ_1 = −1.0		-1.277	0.277	27.7%
		(.346)		
ρ = −0.5		-.521	0.021	4.2%
		(.122)		
		Censored Sample		
β_0 = 0.0	-.435	.013	0.013	—
	(.027)	(.083)		
β_1 = 1.0	.807	1.065	0.065	6.5%
	(.027)	(.054)		
$\sigma_{\varepsilon\varepsilon}$ = 1.0	.916	1.054	0.054	5.4%
	(.029)	(.062)		
γ_0 = 0.0		.021	0.021	—
		(.023)		
γ_1 = −1.0		-1.043	0.043	4.3%
		(.032)		
ρ = −0.5		-.538	0.038	7.6%
		(.078)		

* Truncated sample size.

** Standard errors in parentheses.

There are two problems with Muthén and Jöreskog's results. First, it cannot be identified that whether the standard errors reported in the study are the RMSE's or ASE's. Second, since misspecified models are used in Monte Carlo experiments, it is difficult to distinguish whether the biased results are caused by the truncated sample or by the misspecification.[2]

Comparable Monte Carlo experiment results from a model specified in this study (Chapter 4, Sections 4.1 and 4.2) are presented in Appendix C, Tables C.1.3.A, C.1.3.B, and C.1.3.E.[3]

---

[2] It is showed that a probit model is sensitive to model specification (Yatchew and Griliches, 1985).

[3] Results are interpreted in Chapter 4, Section 4.2.1.

APPENDIX B

NOTATION USED IN REPORTING MONTE CARLO RESULTS

APPENDIX B

NOTATION USED IN REPORTING MONTE CARLO RESULTS[1]

To summarize results from Monte Carlo experiments, let $\hat{\delta}_i$, a kx1 vector, be the estimate of the parameter vector obtained from the i^{th} replication, and $\hat{\pi}_i$ is the corresponding kxk variance-covariance matrix calculated as the inverse of the negative of the second derivatives matrix of the log-likelihood function at the maximum likelihood estimates.[2]

First, the mean estimate of the parameter vector, MEAN, is defined as

$$\text{MEAN} \equiv \bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}\hat{\delta}_i,$$

where N is the total number of replications (N = 500 in this study). A measure of the bias, BIAS, can be defined as

$$\text{BIAS} \equiv \bar{\delta} - \delta^*,$$

where δ^* is the true value of the parameter vector. In addition, define %BIAS by

$$\%\text{BIAS} \equiv \frac{\text{abs}[\ \text{BIAS}\ ]}{\delta^*} \cdot 100\%.$$

---

1 Adapted from Dhrymes (1970), Section 8.6, pp. 372 - 380.

2 In the GAUSS optimization procedure, instead of using an analytical second derivative, a numerical second derivative that is based on an analytical first derivative is used.

Further, define average standard error, ASE, as

$$\text{ASE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\text{diag}[\hat{\pi}_i]}\ ,$$

the covariance matrix about the true parameter value, $\bar{COV}^*(\hat{\delta})$, as

$$\bar{COV}^*(\hat{\delta}) = \frac{1}{N}\sum_{i=1}^{N}(\hat{\delta}_i - \delta^*)(\hat{\delta}_i - \delta^*)',$$

and root mean square errors, RMSE, as

$$\text{RMSE} = \sqrt{\text{diag}\left[\bar{COV}^*(\hat{\delta})\right]}.$$

To examine Monte Carlo experiment results, it is important to check both the RMSE and ASE. Under ideal condition, RMSE and ASE should be very close to each other. The RMSE is very different from ASE if 1) the model is misspecified; 2) the estimator does not lead to reliable parameter estimates; or 3) the variance-covariance matrix cannot be calculated using the regular formula.

For the purpose of comparing efficiency among estimators, define

$$D_{b;j} = \det\left[\bar{COV}^*(\hat{\delta})_{b;j}\right],$$

where b specifies a sub-vector of the parameter vector, j indicates the j^{th} type of estimator and $\bar{COV}^*(\hat{\delta})_{b;j}$ is the corresponding covariance matrix about the true parameter value. For different estimators, if the $D_{b;j}$'s are defined over the same sample, their (relative) magnitudes can be treated as an indicator of "efficiency." For example, $D_{b;i} > D_{b;j}$ indicates that the j^{th} estimator is more efficient than the i^{th} estimator. In this study, although different estimators are based on different data sets,[3] the (relative) magnitude of $D_{b;j}$ can still be treated as an indicator of efficiency.

In Appendix C, MEAN's, BIAS's, %BIAS, RMSE's, ASE's, $D_{b;j}$'s as well as the average log-likelihood value and its standard error (SE),[4] summarize the results of Monte Carlo experiments.

---

[3] In this study, although different estimators are based on different data sets, all the data sets are developed from the same population and contain an identical proportion of respondents to non-respondents. For non-respondents, different data sets contain either the real observations or some statistics estimated from the same population.

[4] These are simply the mean and standard error of the maximum log-likelihood values from the N (= 500) replications.

APPENDIX C

MONTE CARLO EXPERIMENT RESULTS

APPENDIX C

MONTE CARLO EXPERIMENT RESULTS

This appendix presents the results from Monte Carlo experiments for self-selection models with measurement errors with a linear demand, a Tobit demand, and a probit demand equation. Based on different correlation measures between the demand and self-selection equations, three sequences of simulation are conducted for each model (ρ = 0.25, 0.5, and 0.75), and each sequence of simulation has 500 replications.

C.1.1 Estimates from a self-selection model with measurement errors and a linear demand equation (ρ = 0.25)

Results presented below are based on ρ = 0.25.[1] Estimates are obtained after 500 replications, and the average number of respondents is 499.0820 (SE = 21.4049) out of 1,000.

---

[1] Statistics reported in Tables are defined in Appendix B.

Table C.1.1.A Linear demand, OLS estimates without correcting for self-selection, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	6.0065	0.0065	0.11%	0.2328	0.2234
$\beta_1 = 4$	4.0876	0.0876	2.19%	0.1096	0.0640
$\beta_2 = -3$	-3.0028	-0.0028	0.09%	0.0611	0.0586
$\sigma^2 = 1$	0.9868	-0.0132	1.32%	0.0637	

$D_{(\beta,\sigma^2);S} = 6.4531e\text{-}10$

Table C.1.1.B Linear demand, correcting for self-selection bias using censored sample, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5026	0.0026	0.17%	0.2070	0.1954
$\gamma_1 = 1$	1.0120	0.0120	1.20%	0.0794	0.0786
$\gamma_2 = -3$	-3.0259	-0.0259	0.86%	0.1864	0.1799
$\beta_0 = 6$	6.0036	0.0036	0.06%	0.2335	0.2068
$\beta_1 = 4$	4.0014	0.0014	0.04%	0.0831	0.0744
$\beta_2 = -3$	-3.0004	-0.0004	0.01%	0.0612	0.0544
$\sigma^2 = 1$	0.9968	-0.0032	0.32%	0.0642	0.0607
$\rho = 0.25$	0.2490	-0.0010	0.40%	0.1342	0.1221
Average log-likelihood = -928.1916 (SE = 40.6244)					

$D_{(\gamma,\rho);CEN} = 8.8139e\text{-}09$

$D_{(\beta,\sigma^2);CEN} = 3.0124e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 1.1842e\text{-}18$

Table C.1.1.C Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ, ρ = 0.25

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5476	0.0476	3.17%	0.3664	0.3208
γ_1 = 1	1.0338	0.0338	3.38%	0.1420	0.1259
γ_2 = -3	-3.0861	-0.0861	2.87%	0.3651	0.3116
β_0 = 6	6.0038	0.0038	0.06%	0.2336	0.2090
β_1 = 4	4.0013	0.0013	0.03%	0.0843	0.0741
β_2 = -3	-3.0005	-0.0005	0.02%	0.0613	0.0549
σ^2 = 1	0.9972	-0.0028	0.28%	0.0647	0.0602
ρ = 0.25	0.2522	0.0022	0.88%	0.1372	0.1232

Average log-likelihood = -1157.1148 (SE = 41.9968)

$D_{(\gamma,\rho);ME1}$ = 1.3607e-07

$D_{(\beta,\sigma^2);ME1}$ = 3.1814e-10

$D_{(\gamma,\beta,\sigma^2,\rho);ME1}$ = 1.8437e-17

Table C.1.1.D Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, ρ = 0.25

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5130	0.0130	0.87%	0.3520	0.3180
γ_1 = 1	1.0228	0.0228	2.28%	0.1353	0.1241
γ_2 = -3	-3.0538	-0.0538	1.79%	0.3467	0.3079
β_0 = 6	6.0033	0.0033	0.06%	0.2335	0.2094
β_1 = 4	4.0003	0.0003	0.01%	0.0844	0.0744
β_2 = -3	-3.0005	-0.0005	0.02%	0.0613	0.0550
σ^2 = 1	0.9974	-0.0026	0.26%	0.0647	0.0603
ρ = 0.25	0.2516	0.0016	0.64%	0.1368	0.1229

Average log-likelihood = -1157.1736 (SE = 42.0587)

$D_{(\gamma,\rho);ME2}$ = 1.0158e-07

$D_{(\beta,\sigma^2);ME2}$ = 3.1997e-10

$D_{(\gamma,\beta,\sigma^2,\rho);ME2}$ = 1.3851e-17

C.1.2 Estimates from a self-selection model with measurement errors and a linear demand equation (ρ = 0.5)

Results presented below are based on ρ = 0.5. Estimates are obtained after 500 replications, and the average number of respondents is 498.0860 (SE = 22.1867) out of 1,000.

Table C.1.2.A Linear demand, OLS estimates without correcting for self-selection bias, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	5.9969	-0.0031	0.05%	0.2079	0.2209
$\beta_1 = 4$	4.1714	0.1714	4.29%	0.1828	0.0632
$\beta_2 = -3$	-3.0025	-0.0025	0.08%	0.0539	0.0579
$\sigma^2 = 1$	0.9649	-0.0351	3.51%	0.0704	

$D_{(\beta,\sigma^2);S} = 1.4168e\text{-}09$

Table C.1.2.B Linear demand, correcting for self-selection bias using censored sample, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.4988	-0.0012	0.08%	0.2059	0.1955
$\gamma_1 = 1$	1.0116	0.0116	1.16%	0.0790	0.0793
$\gamma_2 = -3$	-3.0231	-0.0231	0.77%	0.1861	0.1818
$\beta_0 = 6$	5.9945	-0.0055	0.09%	0.2060	0.2074
$\beta_1 = 4$	4.0023	0.0023	0.06%	0.0780	0.0732
$\beta_2 = -3$	-2.9989	0.0011	0.04%	0.0532	0.0542
$\sigma^2 = 1$	0.9967	-0.0033	0.33%	0.0659	0.0617
$\rho = 0.5$	0.4946	-0.0054	1.08%	0.1112	0.1083
Average log-likelihood = -914.5157 (SE = 42.3432)					

$D_{(\gamma,\rho);CEN} = 5.4306e\text{-}09$

$D_{(\beta,\sigma^2);CEN} = 1.6633e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 4.0948e\text{-}19$

Table C.1.2.C Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5230	0.0230	1.53%	0.3428	0.3092
$\gamma_1 = 1$	1.0221	0.0221	2.21%	0.1359	0.1222
$\gamma_2 = -3$	-3.0557	-0.0557	1.86%	0.3439	0.3003
$\beta_0 = 6$	5.9940	-0.0060	0.10%	0.2066	0.2092
$\beta_1 = 4$	4.0021	0.0021	0.05%	0.0794	0.0739
$\beta_2 = -3$	-2.9988	0.0012	0.04%	0.0532	0.0546
$\sigma^2 = 1$	0.9972	-0.0028	0.28%	0.0673	0.0629
$\rho = 0.5$	0.4969	-0.0031	0.62%	0.1127	0.1091
Average log-likelihood = -1144.3670 (SE = 43.1100)					

$D_{(\gamma,\rho);ME1} = 6.8753e\text{-}08$

$D_{(\beta,\sigma^2);ME1} = 1.8757e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);ME1} = 5.4879e\text{-}18$

Table C.1.2.D Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.4978	-0.0022	0.15%	0.3246	0.3061
$\gamma_1 = 1$	1.0137	0.0137	1.37%	0.1272	0.1213
$\gamma_2 = -3$	-3.0318	-0.0318	1.06%	0.3195	0.2966
$\beta_0 = 6$	5.9935	-0.0065	0.11%	0.2066	0.2094
$\beta_1 = 4$	4.0007	0.0007	0.02%	0.0794	0.0740
$\beta_2 = -3$	-2.9988	0.0012	0.04%	0.0532	0.0547
$\sigma^2 = 1$	0.9978	-0.0022	0.22%	0.0673	0.0630
$\rho = 0.5$	0.4959	-0.0041	0.82%	0.1124	0.1089
Average log-likelihood = -1144.4764 (SE = 43.2787)					

$D_{(\gamma,\rho);ME2} = 4.8506e\text{-}08$

$D_{(\beta,\sigma^2);ME2} = 1.8872e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);ME2} = 3.8592e\text{-}18$

C.1.3 Estimates from a self-selection model with measurement errors and a linear demand equation (ρ = 0.75)

Results presented below are based on ρ = 0.75. Estimates are obtained after 500 replications, and the average number of respondents is 498.9100 (SE = 20.9709) out of 1,000.

Table C.1.3.A Linear demand, OLS estimates without correcting for self-selection bias, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	6.0257	0.0257	0.43%	0.2187	0.2157
$\beta_1 = 4$	4.2579	0.2579	6.45%	0.2654	0.0619
$\beta_2 = -3$	-3.0109	-0.0109	0.36%	0.0573	0.0565
$\sigma^2 = 1$	0.9227	-0.0773	7.73%	0.0975	

$D_{(\beta,\sigma^2);S} = 3.2229e\text{-}09$

Table C.1.3.B Linear demand, correcting for self-selection bias using censored sample, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5097	0.0097	0.65%	0.1986	0.1824
$\gamma_1 = 1$	1.0097	0.0097	0.97%	0.0794	0.0733
$\gamma_2 = -3$	-3.0271	-0.0271	0.90%	0.1855	0.1734
$\beta_0 = 6$	6.0165	0.0165	0.28%	0.2078	0.1931
$\beta_1 = 4$	4.0013	0.0013	0.03%	0.0691	0.0642
$\beta_2 = -3$	-3.0038	-0.0038	0.13%	0.0534	0.0504
$\sigma^2 = 1$	0.9938	-0.0062	0.62%	0.0673	0.0632
$\rho = 0.75$	0.7521	0.0021	0.28%	0.0657	0.0616
Average log-likelihood = -891.7293 (SE = 38.4693)					

$D_{(\gamma,\rho);CEN} = 1.8304e\text{-}09$

$D_{(\beta,\sigma^2);CEN} = 1.1114e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 1.0369e\text{-}19$

Table C.1.3.C Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5483	0.0483	3.22%	0.3079	0.2718
$\gamma_1 = 1$	1.0305	0.0305	3.05%	0.1266	0.1103
$\gamma_2 = -3$	-3.0846	-0.0846	2.82%	0.3114	0.2723
$\beta_0 = 6$	6.0174	0.0174	0.29%	0.2080	0.1950
$\beta_1 = 4$	4.0035	0.0035	0.09%	0.0710	0.0655
$\beta_2 = -3$	-3.0039	-0.0039	0.13%	0.0533	0.0507
$\sigma^2 = 1$	0.9920	-0.0080	0.80%	0.0714	0.0647
$\rho = 0.75$	0.7540	0.0040	0.53%	0.0675	0.0619
Average log-likelihood = -1120.3568 (SE = 38.5626)					

$D_{(\gamma,\rho);ME1} = 1.8001e\text{-}08$

$D_{(\beta,\sigma^2);ME1} = 1.4288e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);ME1} = 1.0489e\text{-}18$

Table C.1.3.D Linear demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5290	0.0290	1.93%	0.2935	0.2689
$\gamma_1 = 1$	1.0243	0.0243	2.43%	0.1226	0.1092
$\gamma_2 = -3$	-3.0677	-0.0677	2.26%	0.2982	0.2693
$\beta_0 = 6$	6.0166	0.0166	0.28%	0.2077	0.1953
$\beta_1 = 4$	4.0014	0.0014	0.04%	0.0698	0.0657
$\beta_2 = -3$	-3.0038	-0.0038	0.13%	0.0533	0.0508
$\sigma^2 = 1$	0.9932	-0.0068	0.68%	0.0714	0.0648
$\rho = 0.75$	0.7530	0.0030	0.40%	0.0672	0.0620
Average log-likelihood = -1120.4394 (SE = 38.6073)					

$D_{(\gamma,\rho);ME2} = 1.3920e\text{-}08$

$D_{(\beta,\sigma^2);ME2} = 1.3765e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);ME2} = 8.0515e\text{-}19$

Table C.1.3.E Linear demand, correcting for self-selection bias using truncated sample, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5442	0.0442	2.95%	0.6446	0.5437
$\gamma_1 = 1$	1.0538	0.0538	5.38%	0.3605	0.2746
$\gamma_2 = -3$	-3.1487	-0.1487	4.96%	0.7926	0.6842
$\beta_0 = 6$	6.0123	0.0123	0.21%	0.2099	0.1960
$\beta_1 = 4$	3.9913	-0.0087	0.22%	0.1055	0.0966
$\beta_2 = -3$	-3.0041	-0.0041	0.14%	0.0535	0.0503
$\sigma^2 = 1$	0.9967	-0.0033	0.33%	0.0877	0.0808
$\rho = 0.75$	0.7600	0.0100	1.33%	0.1100	0.0989
Average log-likelihood = -665.0761 (SE = 32.2393)					

$D_{(\gamma,\rho);TRU} = 1.9254e\text{-}05$

$D_{(\beta,\sigma^2);TRU} = 9.4964e\text{-}10$

$D_{(\gamma,\beta,\sigma^2,\rho);TRU} = 3.4627e\text{-}15$

C.2.1 Estimates from a self-selection model with measurement errors and a Tobit demand equation (ρ = 0.25)

Results presented below are based on $\rho = 0.25$. Estimates are obtained after 500 replications, and the average number of respondents is 499.9500 (SE = 22.6791) out of 1,000. In the demand equation, the average number of censored q_i's ($q_i = 0$) is 370.4960 (SE = 19.7898), and the average number of uncensored q_i's ($q_i > 0$) is 129.4540 (SE = 11.1605).

Table C.2.1.A Tobit estimates without correcting for self-selection bias, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	5.9277	-0.0723	1.21%	0.4288	0.3878
$\beta_1 = 4$	4.1536	0.1536	3.84%	0.2491	0.1955
$\beta_2 = -3$	-3.0048	-0.0048	0.16%	0.1523	0.1461
$\sigma^2 = 1$	0.9774	-0.0226	2.26%	0.1257	0.1221
Average log-likelihood = -226.2562 (SE = 19.4390)					

$D_{(\beta,\sigma^2);S} = 1.7151\text{e-}07$

Table C.2.1.B Tobit demand, correcting for self-selection bias using censored sample, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5215	0.0215	1.43%	0.2130	0.2140
$\gamma_1 = 1$	1.0013	0.0013	0.13%	0.0827	0.0836
$\gamma_2 = -3$	-3.0185	-0.0185	0.62%	0.1978	0.1953
$\beta_0 = 6$	5.9834	-0.0166	0.28%	0.4005	0.3910
$\beta_1 = 4$	4.0039	0.0039	0.10%	0.2130	0.2221
$\beta_2 = -3$	-2.9974	0.0026	0.09%	0.1472	0.1462
$\sigma^2 = 1$	0.9988	-0.0012	0.12%	0.1329	0.1289
$\rho = 0.25$	0.2426	-0.0074	2.96%	0.1873	0.1815
Average log-likelihood = -451.5493 (SE = 28.6113)					

$D_{(\gamma,\rho);CEN} = 2.0024\text{e-}08$

$D_{(\beta,\sigma^2);CEN} = 1.0154\text{e-}07$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 7.4238\text{e-}16$

Table C.2.1.C Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5702	0.0702	4.68%	0.3603	0.3481
$\gamma_1 = 1$	1.0175	0.0175	1.75%	0.1316	0.1343
$\gamma_2 = -3$	-3.0711	-0.0711	2.37%	0.3338	0.3328
$\beta_0 = 6$	5.9792	-0.0208	0.35%	0.4093	0.3985
$\beta_1 = 4$	4.0033	0.0033	0.08%	0.2140	0.2255
$\beta_2 = -3$	-2.9954	0.0046	0.15%	0.1499	0.1492
$\sigma^2 = 1$	0.9982	-0.0018	0.18%	0.1341	0.1295
$\rho = 0.25$	0.2427	-0.0073	2.92%	0.1911	0.1837
Average log-likelihood = -679.4912 (SE = 30.5418)					

$D_{(\gamma,\rho);ME1} = 2.4916e\text{-}07$

$D_{(\beta,\sigma^2);ME1} = 1.0763e\text{-}07$

$D_{(\gamma,\beta,\sigma^2,\rho);ME1} = 9.7833e\text{-}15$

Table C.2.1.D Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5557	0.0557	3.71%	0.3530	0.3414
$\gamma_1 = 1$	1.0123	0.0123	1.23%	0.1300	0.1319
$\gamma_2 = -3$	-3.0597	-0.0597	1.99%	0.3265	0.3256
$\beta_0 = 6$	5.9786	-0.0214	0.36%	0.4104	0.3960
$\beta_1 = 4$	4.0029	0.0029	0.07%	0.2137	0.2258
$\beta_2 = -3$	-2.9957	0.0043	0.14%	0.1503	0.1478
$\sigma^2 = 1$	0.9984	-0.0016	0.16%	0.1344	0.1295
$\rho = 0.25$	0.2424	-0.0076	3.04%	0.1909	0.1855
Average log-likelihood = -679.5947 (SE = 30.5940)					

$D_{(\gamma,\rho);ME2} = 1.9532e\text{-}07$

$D_{(\beta,\sigma^2);ME2} = 1.0759e\text{-}07$

$D_{(\gamma,\beta,\sigma^2,\rho);ME2} = 7.5680e\text{-}15$

C.2.2 Estimates from a self-selection model with measurement errors and a Tobit demand equation ($\rho = 0.5$)

Results presented below are based on $\rho = 0.5$. Estimates are obtained after 500 replications, and the average number of respondents is 499.5380 (SE = 21.1736) out of 1,000. In the demand equation, the average number of censored q_i's ($q_i = 0$) is 364.3780 (SE = 18.4454), and the average number of uncensored q_i's ($q_i > 0$) is 135.1600 (SE = 11.7430).

Table C.2.2.A Tobit estimates without correcting for self-selection bias, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	5.8824	-0.1176	1.96%	0.4033	0.3732
$\beta_1 = 4$	4.3040	0.3040	7.60%	0.3626	0.1895
$\beta_2 = -3$	-3.0145	-0.0145	0.48%	0.1471	0.1391
$\sigma^2 = 1$	0.9345	-0.0655	6.55%	0.1348	0.1139
Average log-likelihood = -229.8972 (SE = 20.0318)					

$D_{(\beta,\sigma^2);S} = 4.6211e\text{-}07$

Table C.2.2.B Tobit demand, correcting for self-selection bias using censored sample, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5329	0.0329	2.19%	0.2120	0.2144
$\gamma_1 = 1$	1.0111	0.0111	1.11%	0.0879	0.0837
$\gamma_2 = -3$	-3.0414	-0.0414	1.38%	0.2041	0.1962
$\beta_0 = 6$	5.9847	-0.0153	0.26%	0.3724	0.3715
$\beta_1 = 4$	4.0087	0.0087	0.22%	0.2043	0.2039
$\beta_2 = -3$	-2.9957	0.0043	0.14%	0.1380	0.1365
$\sigma^2 = 1$	0.9854	-0.0146	1.46%	0.1353	0.1272
$\rho = 0.5$	0.4836	-0.0164	3.28%	0.1594	0.1512
Average log-likelihood = -450.5322 (SE = 29.5257)					

$D_{(\gamma,\rho);CEN} = 1.4824e\text{-}08$

$D_{(\beta,\sigma^2);CEN} = 6.4816e\text{-}08$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 3.3847e\text{-}16$

Table C.2.2.C Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ, ρ = 0.5

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5511	0.0511	3.41%	0.3635	0.3454
γ_1 = 1	1.0282	0.0282	2.82%	0.1353	0.1381
γ_2 = -3	-3.0788	-0.0788	2.63%	0.3515	0.3407
β_0 = 6	5.9824	-0.0176	0.29%	0.3758	0.3702
β_1 = 4	4.0106	0.0106	0.27%	0.2066	0.2053
β_2 = -3	-2.9958	0.0042	0.14%	0.1385	0.1355
σ^2 = 1	0.9859	-0.0141	1.41%	0.1354	0.1291
ρ = 0.5	0.4859	-0.0141	2.82%	0.1599	0.1534

Average log-likelihood = -680.7637 (SE = 30.1853)

$D_{(\gamma,\rho);ME1}$ = 1.7301e-07

$D_{(\beta,\sigma^2);ME1}$ = 7.1621e-08

$D_{(\gamma,\beta,\sigma^2,\rho);ME1}$ = 4.3018e-15

Table C.2.2.D Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, ρ = 0.5

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5252	0.0252	1.68%	0.3447	0.3381
γ_1 = 1	1.0205	0.0205	2.05%	0.1285	0.1314
γ_2 = -3	-3.0565	-0.0565	1.88%	0.3282	0.3241
β_0 = 6	5.9809	-0.0191	0.32%	0.3779	0.3765
β_1 = 4	4.0100	0.0100	0.25%	0.2077	0.2081
β_2 = -3	-2.9956	0.0044	0.15%	0.1396	0.1387
σ^2 = 1	0.9862	-0.0138	1.38%	0.1361	0.1289
ρ = 0.5	0.4847	-0.0153	3.06%	0.1607	0.1534

Average log-likelihood = -680.8440 (SE = 30.2236)

$D_{(\gamma,\rho);ME2}$ = 1.2593e-07

$D_{(\beta,\sigma^2);ME2}$ = 7.1884e-08

$D_{(\gamma,\beta,\sigma^2,\rho);ME2}$ = 3.1359e-15

C.2.3 Estimates from a self-selection model with measurement errors and a Tobit demand equation (ρ = 0.75)

Results presented below are based on $\rho = 0.75$. Estimates are obtained after 500 replications, and the average number of respondents is 499.9580 (SE = 21.8165) out of 1,000. In the demand equation, the average number of censored q_i's ($q_i = 0$) is 361.5160 (SE = 18.9821), and the average number of uncensored q_i's ($q_i > 0$) is 138.4420 (SE = 11.1505).

Table C.2.3.A Tobit estimates without correcting for self-selection bias, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	5.8553	-0.1447	2.41%	0.3971	0.3561
$\beta_1 = 4$	4.4434	0.4434	11.09%	0.4808	0.1848
$\beta_2 = -3$	-3.0273	-0.0273	0.91%	0.1383	0.1328
$\sigma^2 = 1$	0.8767	-0.1233	12.33%	0.1633	0.1054
Average log-likelihood = -228.8807 (SE = 18.4061)					

$D_{(\beta,\sigma^2);S} = 6.5670e\text{-}07$

Table C.2.3.B Tobit demand, correcting for self-selection bias using censored sample, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.4990	-0.0010	0.07%	0.1959	0.2091
$\gamma_1 = 1$	1.0177	0.0177	1.77%	0.0835	0.0821
$\gamma_2 = -3$	-3.0368	-0.0368	1.23%	0.1862	0.1943
$\beta_0 = 6$	6.0087	0.0087	0.15%	0.3409	0.3408
$\beta_1 = 4$	4.0121	0.0121	0.30%	0.1780	0.1804
$\beta_2 = -3$	-3.0057	-0.0057	0.19%	0.1254	0.1232
$\sigma^2 = 1$	0.9882	-0.0118	1.18%	0.1302	0.1280
$\rho = 0.75$	0.7396	-0.0104	1.39%	0.1028	0.0973
Average log-likelihood = -443.3619 (SE = 26.9542)					

$D_{(\gamma,\rho);CEN} = 4.8510e\text{-}09$

$D_{(\beta,\sigma^2);CEN} = 2.8632e\text{-}08$

$D_{(\gamma,\beta,\sigma^2,\rho);CEN} = 5.7769e\text{-}17$

Table C.2.3.C Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ, ρ = 0.75

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5645	0.0645	4.30%	0.3311	0.3295
γ_1 = 1	1.0362	0.0362	3.62%	0.1391	0.1279
γ_2 = -3	-3.1039	-0.1039	3.46%	0.3510	0.3210
β_0 = 6	6.0113	0.0113	0.19%	0.3453	0.3421
β_1 = 4	4.0101	0.0101	0.25%	0.1813	0.1839
β_2 = -3	-3.0044	-0.0044	0.15%	0.1260	0.1229
σ^2 = 1	0.9858	-0.0142	1.42%	0.1325	0.1305
ρ = 0.75	0.7401	-0.0099	1.32%	0.1058	0.0979
Average log-likelihood = -673.0108 (SE = 29.1919)					

$D_{(\gamma,\rho);ME1}$ = 5.6626e-08

$D_{(\beta,\sigma^2);ME1}$ = 3.3398e-08

$D_{(\gamma,\beta,\sigma^2,\rho);ME1}$ = 6.8714e-16

Table C.2.3.D Tobit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, ρ = 0.75

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5348	0.0348	2.32%	0.3169	0.3248
γ_1 = 1	1.0260	0.0260	2.60%	0.1330	0.1262
γ_2 = -3	-3.0763	-0.0763	2.54%	0.3360	0.3173
β_0 = 6	6.0116	0.0116	0.19%	0.3454	0.3425
β_1 = 4	4.0068	0.0068	0.17%	0.1806	0.1841
β_2 = -3	-3.0050	-0.0050	0.17%	0.1261	0.1233
σ^2 = 1	0.9891	-0.0109	1.09%	0.1346	0.1324
ρ = 0.75	0.7397	-0.0103	1.37%	0.1060	0.0981
Average log-likelihood = -673.0461 (SE = 29.1821)					

$D_{(\gamma,\rho);ME2}$ = 4.1771e-08

$D_{(\beta,\sigma^2);ME2}$ = 3.3442e-08

$D_{(\gamma,\beta,\sigma^2,\rho);ME2}$ = 5.1214e-16

C.3.1 Estimates from a self-selection model with measurement errors and a probit demand equation ($\rho = 0.25$)

Results presented below are based on $\rho = 0.25$. Estimates are obtained after 500 replications, and the average number of respondents is 502.3760 (SE = 20.6115) out of 1,000. In the demand equation, the average number of left censored q_i's ($q_i = 0$) is 372.0040 (SE = 18.5927), and the average number of right censored q_i's ($q_i = 1$) is 130.3720 (SE = 11.2154).

Table C.3.1.A Probit estimates without correcting for self-selection bias, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	6.2406	0.2406	4.01%	0.9353	0.8337
$\beta_1 = 4$	4.2972	0.2972	7.43%	0.5777	0.4633
$\beta_2 = -3$	-3.1363	-0.1363	4.54%	0.3961	0.3442
Average log-likelihood = -85.3713 (SE = 10.3680)					

$D_{\beta;S} = 1.9852\text{e-}04$

Table C.3.1.B Probit demand, correcting for self-selection bias using censored sample, $\rho = 0.25$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5329	0.0329	2.19%	0.2293	0.2187
$\gamma_1 = 1$	1.0144	0.0144	1.44%	0.0849	0.0852
$\gamma_2 = -3$	-3.0501	-0.0501	1.67%	0.2122	0.2005
$\beta_0 = 6$	6.1967	0.1967	3.28%	0.9208	0.8351
$\beta_1 = 4$	4.1095	0.1095	2.74%	0.5539	0.5104
$\beta_2 = -3$	-3.0892	-0.0892	2.97%	0.3863	0.3471
$\rho = 0.25$	0.2527	0.0027	1.08%	0.3060	0.2918
Average log-likelihood = -308.9247 (SE = 20.4965)					

$D_{(\gamma,\rho);CEN} = 6.4142\text{e-}08$

$D_{\beta;CEN} = 2.1815\text{e-}04$

$D_{(\gamma,\beta,\rho);CEN} = 5.7177\text{e-}12$

Table C.3.1.C Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ, ρ = 0.25

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.6115	0.1115	7.43%	0.3756	0.3605
γ_1 = 1	1.0449	0.0449	4.49%	0.1473	0.1384
γ_2 = -3	-3.1454	-0.1454	4.85%	0.3698	0.3477
β_0 = 6	6.2021	0.2021	3.37%	0.9239	0.8415
β_1 = 4	4.1132	0.1132	2.83%	0.5560	0.5124
β_2 = -3	-3.0912	-0.0912	3.04%	0.3877	0.3490
ρ = 0.25	0.2534	0.0034	1.36%	0.3092	0.2964
Average log-likelihood = -537.6684 (SE = 22.0822)					

$D_{(\gamma,\rho);ME1}$ = 8.7102e-07

$D_{\beta;ME1}$ = 2.2009e-04

$D_{(\gamma,\beta,\rho);ME1}$ = 7.8470e-11

Table C.3.1.D Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, ρ = 0.25

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
γ_0 = 1.5	1.5746	0.0746	4.97%	0.3539	0.3588
γ_1 = 1	1.0335	0.0335	3.35%	0.1382	0.1381
γ_2 = -3	-3.1112	-0.1112	3.71%	0.3438	0.3468
β_0 = 6	6.2002	0.2002	3.34%	0.9230	0.8498
β_1 = 4	4.1111	0.1111	2.78%	0.5554	0.5160
β_2 = -3	-3.0904	-0.0904	3.01%	0.3873	0.3539
ρ = 0.25	0.2529	0.0029	1.16%	0.3077	0.2922
Average log-likelihood = -537.7948 (SE = 22.0741)					

$D_{(\gamma,\rho);ME2}$ = 5.7620e-07

$D_{\beta;ME2}$ = 2.2082e-04

$D_{(\gamma,\beta,\rho);ME2}$ = 5.2055e-11

C.3.2 Estimates from a self-selection model with measurement errors and a probit demand equation ($\rho = 0.5$)

Results presented below are based on $\rho = 0.5$. Estimates are obtained after 500 replications, and the average number of respondents is 501.0820 (SE = 23.3509) out of 1,000. In the demand equation, the average number of left censored q_i's ($q_i = 0$) is 366.9400 (SE = 20.4881), and the average number of right censored q_i's ($q_i = 1$) is 134.1420 (SE = 11.8349).

Table C.3.2.A Probit estimates without correcting for self-selection bias, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	6.2315	0.2315	3.86%	0.8671	0.8428
$\beta_1 = 4$	4.5082	0.5082	12.71%	0.7114	0.4876
$\beta_2 = -3$	-3.1767	-0.1767	5.89%	0.3926	0.3519
Average log-likelihood = -83.3600 (SE = 10.2106)					

$D_{\beta;S} = 4.7665e\text{-}04$

Table C.3.2.B Probit demand, correcting for self-selection bias using censored sample, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5067	0.0067	0.45%	0.2121	0.2194
$\gamma_1 = 1$	1.0173	0.0173	1.73%	0.0847	0.0864
$\gamma_2 = -3$	-3.0376	-0.0376	1.25%	0.1977	0.2034
$\beta_0 = 6$	6.1576	0.1576	2.63%	0.8458	0.8762
$\beta_1 = 4$	4.1689	0.1689	4.22%	0.5658	0.5607
$\beta_2 = -3$	-3.0930	-0.0930	3.10%	0.3672	0.3718
$\rho = 0.5$	0.4529	-0.0471	9.42%	0.2570	0.2719
Average log-likelihood = -306.9455 (SE = 19.9936)					

$D_{(\gamma,\rho);CEN} = 3.3612e\text{-}08$

$D_{\beta;CEN} = 1.8584e\text{-}04$

$D_{(\gamma,\beta,\rho);CEN} = 2.6645e\text{-}12$

Table C.3.2.C Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5762	0.0762	5.08%	0.3800	0.3657
$\gamma_1 = 1$	1.0346	0.0346	3.46%	0.1470	0.1400
$\gamma_2 = -3$	-3.1046	-0.1046	3.49%	0.3794	0.3516
$\beta_0 = 6$	6.1572	0.1572	2.62%	0.8451	0.8649
$\beta_1 = 4$	4.1653	0.1653	4.13%	0.5639	0.5706
$\beta_2 = -3$	-3.0917	-0.0917	3.06%	0.3661	0.3706
$\rho = 0.5$	0.4594	-0.0406	8.12%	0.2606	0.2810
Average log-likelihood = -536.5940 (SE = 22.4278)					

$D_{(\gamma,\rho);ME1} = 5.4912e\text{-}07$

$D_{\beta;ME1} = 1.8759e\text{-}04$

$D_{(\gamma,\beta,\rho);ME1} = 4.3700e\text{-}11$

Table C.3.2.D Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, $\rho = 0.5$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5398	0.0398	2.65%	0.3624	0.3494
$\gamma_1 = 1$	1.0226	0.0226	2.26%	0.1368	0.1378
$\gamma_2 = -3$	-3.0695	-0.0695	2.32%	0.3529	0.3432
$\beta_0 = 6$	6.1533	0.1533	2.56%	0.8458	0.8823
$\beta_1 = 4$	4.1607	0.1607	4.02%	0.5634	0.5731
$\beta_2 = -3$	-3.0900	-0.0900	3.00%	0.3662	0.3764
$\rho = 0.5$	0.4586	-0.0414	8.28%	0.2591	0.2795
Average log-likelihood = -536.7139 (SE = 22.4111)					

$D_{(\gamma,\rho);ME2} = 3.6906e\text{-}07$

$D_{\beta;ME2} = 1.8893e\text{-}04$

$D_{(\gamma,\beta,\rho);ME2} = 2.9529e\text{-}11$

C.3.3 Estimates from a self-selection model with measurement errors and a probit demand equation (ρ = 0.75)

Results presented below are based on $\rho = 0.75$. Estimates are obtained after 500 replications, and the average number of respondents is 501.5800 (SE = 22.6946) out of 1,000. In the demand equation, the average number of left censored q_i's ($q_i = 0$) is 362.0720 (SE = 19.2150), and the average number of right censored q_i's ($q_i = 1$) is 139.5080 (SE = 11.6935).

Table C.3.3.A Probit estimates without correcting for self-selection bias, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\beta_0 = 6$	6.3941	0.3941	6.57%	0.9873	0.8639
$\beta_1 = 4$	4.7338	0.7338	18.35%	0.9147	0.5130
$\beta_2 = -3$	-3.2647	-0.2647	8.82%	0.4696	0.3633
Average log-likelihood = -81.9309 (SE = 10.5347)					

$D_{\beta;s} = 9.4296\text{e-}04$

Table C.3.3.B Probit demand, correcting for self-selection bias using censored sample, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5155	0.0155	1.03%	0.2221	0.2164
$\gamma_1 = 1$	1.0151	0.0151	1.51%	0.0842	0.0858
$\gamma_2 = -3$	-3.0394	-0.0394	1.31%	0.2093	0.2023
$\beta_0 = 6$	6.2153	0.2153	3.59%	0.9112	0.8915
$\beta_1 = 4$	4.1852	0.1852	4.63%	0.6221	0.5865
$\beta_2 = -3$	-3.1079	-0.1079	3.60%	0.4056	0.3839
$\rho = 0.75$	0.6729	-0.0771	10.28%	0.2064	0.2232
Average log-likelihood = -304.5480 (SE = 19.9643)					

$D_{(\gamma,\rho);CEN} = 4.2098\text{e-}08$
$D_{\beta;CEN} = 1.3845\text{e-}04$
$D_{(\gamma,\beta,\rho);CEN} = 4.7759\text{e-}12$

Table C.3.3.C Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5747	0.0747	4.98%	0.3503	0.3566
$\gamma_1 = 1$	1.0448	0.0448	4.48%	0.1481	0.1385
$\gamma_2 = -3$	-3.1198	-0.1198	3.99%	0.3702	0.3475
$\beta_0 = 6$	6.2255	0.2255	0.04%	0.9136	0.8503
$\beta_1 = 4$	4.1936	0.1936	4.84%	0.6260	0.5624
$\beta_2 = -3$	-3.1121	-0.1121	3.74%	0.4071	0.3655
$\rho = 0.75$	0.6772	-0.0728	9.71%	0.2047	0.2203

Average log-likelihood = -532.9000 (SE = 22.0023)

$D_{(\gamma,\rho);ME1} = 5.3192e\text{-}07$

$D_{\beta;ME1} = 1.4048e\text{-}04$

$D_{(\gamma,\beta,\rho);ME1} = 5.8446e\text{-}11$

Table C.3.3.D Probit demand, correcting for self-selection using measurement errors model with μ_i and Σ_i, $\rho = 0.75$

Parameter	MEAN	BIAS	%BIAS	RMSE	ASE
$\gamma_0 = 1.5$	1.5379	0.0379	2.53%	0.3256	0.3475
$\gamma_1 = 1$	1.0333	0.0333	3.30%	0.1392	0.1355
$\gamma_2 = -3$	-3.0855	-0.0855	2.85%	0.3414	0.3384
$\beta_0 = 6$	6.2195	0.2195	3.66%	0.9117	0.8446
$\beta_1 = 4$	4.1877	0.1877	4.69%	0.6232	0.5606
$\beta_2 = -3$	-3.1095	-0.1095	3.65%	0.4060	0.3628
$\rho = 0.75$	0.6751	-0.0749	9.99%	0.2063	0.2202

Average log-likelihood = -532.9650 (SE = 21.9812)

$D_{(\gamma,\rho);ME2} = 3.6802e\text{-}07$

$D_{\beta;ME2} = 1.3780e\text{-}04$

$D_{(\gamma,\beta,\rho);ME2} = 4.1849e\text{-}11$

APPENDIX D

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A LINEAR DEMAND EQUATION

APPENDIX D

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A LINEAR DEMAND EQUATION

This GAUSS program is used to conduct Monte Carlo experiments for the self-selection model with measurement errors and a linear demand equation (ρ = 0.25).

```
new;
use optmum;
output file = tl.out;

iter = 1;
do while iter le 500;

recal:

@--
npop: population size
nx: number of variables (x's and e's)
nb: number of blocks (nb)
bobs: number of obs in a block
ss: number of obs used to calculate variance-covariance
smp: random sub-sample / population (0 <= smp <= 1)
--@
npop = 10000;
nx = 5;
nb = 250;
bobs = npop / nb;
ss = 200;
smp = 0.1;


@-- draw random sample as population --@
x = rndn(npop,nx);
```

```
let v[5,5] = 1.44   0.24  0.096  0   0
             0.24   1     0.24   0   0
             0.096  0.24  0.64   0   0
             0      0     0      1   0.25
             0      0     0      0.25 1;

sqrtv = chol(v);
x = x * sqrtv;

let a[1,5] = 3 1.5 4 0 0;
x = x + a;

evar = vcx(x[1:ss,1 2]);
clear v, sqrtv, a;


@-- divide population into blocks --@
si = 2000;
s = npop / si;
sg = nb / s;
w = eye(sg) .*. ones(bobs,1);
sx = x[1:si,1 2];
msx = (sx'w/bobs)' .*. ones(bobs,1);
vsx = ((((sx-msx)^2)
     ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
     w / (bobs-1) )' .*. ones(bobs,1);
xx = msx~vsx;

i = 2;
do while i <= s;
    l = si * (i - 1) + 1;
    h = si * i;
    sx = x[l:h,1 2];
    msx = (sx'w/bobs)' .*. ones(bobs,1);
    vsx = ((((sx-msx)^2)
         ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
         w / (bobs-1) )' .*. ones(bobs,1);
    sx = msx~vsx;
    xx = xx|sx;
    i = i + 1;
endo;

clear i, l, h, msx, vsx, sx;

@-- xx = [x1 x2 x3 e1 e2 m1 m2 v11 v22 v21] --@

x = x~xx;

clear xx;


@--
extract a random sub-sample from population xx,
```

```
and variables in the random sub-sample are
[x1 x2 x3 e1 e2 m1 m2 v11 v22 v21].
--@
i = rndu(npop,1);
x = i~x;
data = selif(x,x[.,1] .le smp);
clear x;
data = data[.,2:(cols(data))];
obs = rows(data);


@--
generate dependent variables S* and D*
S* = 1.5 + 1 * x1 - 3 * x2 + e1
D* = 6 + 4 * x2 - 3 * x3 + e2
--@
let bs[3,1] = 1.5  1  -3;
let bd[3,1] = 6  4  -3;
s = (ones(obs,1)~data[.,1 2]) * bs + data[.,4];
d = (ones(obs,1)~data[.,2 3]) * bd + data[.,5];


@--
rearrange data according to
s > 0: xyes = [d x1 x2 x3], and
s <= 0: xno = [m1 m2 v11 v22 v21]
also data = [s d x1 x2 x3 m1 m2 v11 v22 v21].
--@
data = s~d~data[.,1:3 6:10];
clear s, d;

yes = selif(data, data[.,1] .gt 0);
no = selif(data, data[.,1] .le 0);
clear data;
nyes = rows(yes);
nno = rows(no);

sxyes = ones(nyes,1)~yes[.,3 4];
dxyes = ones(nyes,1)~yes[.,4 5];
d = yes[.,2];
clear yes;

sxno = ones(nno,1)~no[.,3 4];
sxnom = ones(nno,1)~no[.,6 7];
sxnov = no[.,8 9 10];
clear no;


@--
starting values for
[s0 s1 s2 d0 d1 d2 sigma^2 rho]
--@
b0 = bs|bd|1|0.25;
```

```
@-- gradient tolerance (default = 1e-5) --@
@-- _opgtol = 1e-10; --@

@-- OLS --@
beta = invpd(dxyes'dxyes) * dxyes'd;
s = sumc( (d - dxyes * beta)^2 ) / nyes;
cov = s * invpd(dxyes'dxyes);
stder = sqrt(diag(cov));
output on;
iter~nyes~beta'~stder'~s;
output off;


@-- call and print optmum --@

optset;
_opgdprc = &foc1;
{beta1, f1, g, retcode} = optmum(&fn1, b0);
if retcode ne 0;
    goto recal;
endif;
cov1 = _opfhess;
stder1 = sqrt(diag(cov1));
output on;
(-f1)~beta1'~stder1';
output off;

optset;
_opgdprc = &foc3;
{beta3, f3, g, retcode} = optmum(&fn3, b0);
if retcode ne 0;
    goto recal;
endif;
cov3 = _opfhess;
stder3 = sqrt(diag(cov3));
output on;
(-f3)~beta3'~stder3';
output off;

optset;
_opgdprc = &foc4;
{beta4, f4, g, retcode} = optmum(&fn4, b0);
if retcode ne 0;
    goto recal;
endif;
cov4 = _opfhess;
stder4 = sqrt(diag(cov4));
output on;
(-f4)~beta4'~stder4';
print " ";
output off;
```

```
@-- procedures --@

@--
-(log-likelihood) function for a self-selection model
with measurement errors and a linear demand function
using censored data (non-respondents' independent
variables are observed)
--@
proc fn1(para);
    local bbs, bbd, sigma, rho, dyes, kyes, kno, k;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];

    dyes = d - dxyes * bbd;
    kyes = (1 / sigma) * pdfn(dyes / sigma)
        .* cdfnc( (- sxyes * bbs - rho * dyes
                  / sigma)
                / sqrt(1 - rho^2) );

    kno = cdfn(- sxno * bbs);

    k = ln(kyes|kno);

    retp(-sumc(k));
endp;


proc foc1(para);
    local bbs, bbd, sigma, rho, f, cyes,
        gp, gc, cno, hp, hc, fbbd, fss,
        gbbd, gss, gbbs, grho, hbbs, k;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    f = (1 / sigma) * pdfn( (d - dxyes * bbd) / sigma);
    cyes = ( - sxyes * bbs
           - rho * (d - dxyes *bbd) / sigma )
         / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes);
    gc = cdfnc(cyes);
    hp = pdfn( - sxno * bbs);
    hc = cdfn( - sxno * bbs);

    fbbd = sumc(
          ((d - dxyes * bbd) / sigma^2) .* dxyes  );
    fss = sumc(
         ((d - dxyes * bbd)^2 - sigma^2)
           / (2 * sigma^4)    );
```

```
    gbbd = sumc(
        (- gp ./ gc) * (rho / sigma) .* dxyes );
    gss = sumc(
        (- gp ./ gc) .* (rho * (d - dxyes * bbd))
        / (2 * sigma^3) );
    gbbs = sumc( ( gp ./ gc) .* sxyes );
    grho = sumc( (- gp ./ gc )
        .* ( -((d - dxyes * bbd) / sigma)
        + rho * ( - sxyes * bbs - rho * (d - dxyes * bbd)
        / sigma) / (1 - rho^2) ) );
    hbbs = sumc( (hp ./ hc) .* (- sxno) );

    k = (gbbs'+hbbs') ~ (fbbd + gbbd)'
        ~ (fss + gss) ~ grho;

    retp(-k);
endp;


@--
-(log-likelihood) function for a self-selection model
with measurement errors and a linear demand function
using empirical block mean and estimated variance-
covariance (200 obs. from the population)
for measurement error
--@
proc fn3(para);
    local bbs, bbd, sigma, rho, dyes, kyes,
        bbsn, delta, kno, k;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];

    dyes = d - dxyes * bbd;
    kyes = (1 / sigma) * pdfn(dyes / sigma)
        .* cdfnc( (- sxyes * bbs - rho * dyes
                / sigma)
              / sqrt(1 - rho^2) );

    bbsn = bbs[2 3,.];

    delta = sqrt(1 + bbsn'evar*bbsn);

    kno = cdfn(- sxnom * bbs / delta);

    k = ln(kyes|kno);

    retp(-sumc(k));
endp;
```

```
proc foc3(para);
    local bbs, b1, b2, b3, bbd, sigma, rho, f, cyes,
        gp, gc, cno, hp, hc, h, fbbd, fss,
        gbbd, gss, gbbs, grho, hb1, hb2, hb3, k;

    bbs = para[1 2 3,.];
    b1 = bbs[1,.];
    b2 = bbs[2,.];
    b3 = bbs[3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    f = (1 / sigma) * pdfn( (d - dxyes * bbd) / sigma);
    cyes = ( - sxyes * bbs
            - rho * (d - dxyes *bbd) / sigma )
          / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes);
    gc = cdfnc(cyes);
    cno = (- sxnom * bbs) ./
        sqrt(1 + b2^2 * evar[1,1] + b3^2 * evar[2,2]
            + 2 * b2 * b3 * evar[1,2]);
    hp = (1 / sqrt(1 + b2^2 * evar[1,1]
                + b3^2 * evar[2,2]
                + 2 * b2 * b3 * evar[1,2]))
        .* pdfn(cno);
    hc = cdfn(cno);
    h =  1 + b2^2 * evar[1,1]
        + b3^2 * evar[2,2] + 2 * b2 * b3 * evar[1,2];

    fbbd = sumc(
        ((d - dxyes * bbd) / sigma^2) .* dxyes  );
    fss = sumc(
        ((d - dxyes * bbd)^2 - sigma^2)
         / (2 * sigma^4)    );
    gbbd = sumc(
        (- gp ./ gc) * (rho / sigma) .* dxyes  );
    gss = sumc(
        (- gp ./ gc) .* (rho * (d - dxyes * bbd))
        / (2 * sigma^3)   );
    gbbs = sumc(  ( gp ./ gc) .*  sxyes   );
    grho = sumc(  (- gp ./ gc )
        .* ( -((d - dxyes * bbd) / sigma)
        + rho * ( - sxyes * bbs - rho * (d - dxyes * bbd)
        / sigma) / (1 - rho^2) )  );
    hb1 = sumc( - hp ./ hc  );
    hb2 = sumc(  (hp ./ hc) .* ( - sxnom[.,2]
        + sxnom * bbs .* (2 * b2 * evar[1,1]
        + 2 * b3 * evar[1,2]) ./ (2 * h ) )  );
    hb3 = sumc(  (hp ./ hc) .* ( - sxnom[.,3]
        + sxnom * bbs .* (2 * b3 * evar[2,2]
        + 2 * b2 * evar[1,2]) ./ (2 * h ) )  );

k = (gbbs'+(hb1~hb2~hb3)) ~ (fbbd + gbbd)'
```

```
        ~ (fss + gss) ~ grho;

    retp(-k);
endp;


@--
-(log-likelihood) function for a self-selection model
with measurement errors and a linear demand function
using empirical block mean and empirical block
variance-covariance for measurement error
--@
proc fn4(para);
    local bbs, bbd, sigma, rho, dyes, kyes,
          bbs1, bbs2, delta, kno, k;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];

    dyes = d - dxyes * bbd;
    kyes = (1 / sigma) * pdfn(dyes / sigma)
        .* cdfnc( (- sxyes * bbs - rho * dyes
                  / sigma)
                / sqrt(1 - rho^2) );

    bbs1 = bbs[2,.];
    bbs2 = bbs[3,.];
    delta = sqrt(1 + bbs1^2 * sxnov[.,1]
                + bbs2^2 * sxnov[.,2]
                + 2 * bbs1 * bbs2 * sxnov[.,3]);

    kno = cdfn(- sxnom * bbs ./ delta);

    k = ln(kyes|kno);

    retp(-sumc(k));
endp;


proc foc4(para);
    local bbs, b1, b2, b3, bbd, sigma, rho, f, cyes,
          gp, gc, cno, hp, hc, h, fbbd, fss,
          gbbd, gss, gbbs, grho, hb1, hb2, hb3, k;

    bbs = para[1 2 3,.];
    b1 = bbs[1,.];
    b2 = bbs[2,.];
    b3 = bbs[3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
```

```
    f = (1 / sigma) * pdfn( (d - dxyes * bbd) / sigma);
    cyes = ( - sxyes * bbs
            - rho * (d - dxyes *bbd) / sigma )
          / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes);
    gc = cdfnc(cyes);
    cno = (- sxnom * bbs) ./
         sqrt(1 + b2^2 * sxnov[.,1] + b3^2 * sxnov[.,2]
              + 2 * b2 * b3 * sxnov[.,3]);
    hp = (1 / sqrt(1 + b2^2 * sxnov[.,1]
                + b3^2 * sxnov[.,2]
                + 2 * b2 * b3 * sxnov[.,3]))
        .* pdfn(cno);
    hc = cdfn(cno);
    h =  1 + b2^2 * sxnov[.,1]
       + b3^2 * sxnov[.,2] + 2 * b2 * b3 * sxnov[.,3];

    fbbd = sumc(
         ((d - dxyes * bbd) / sigma^2) .* dxyes  );
    fss = sumc(
        ((d - dxyes * bbd)^2 - sigma^2)
         / (2 * sigma^4)    );
    gbbd = sumc(
        (- gp ./ gc) * (rho / sigma) .* dxyes  );
    gss = sumc(
        (- gp ./ gc) .* (rho * (d - dxyes * bbd))
        / (2 * sigma^3)   );
    gbbs = sumc(  ( gp ./ gc) .*  sxyes   );
    grho = sumc(   (- gp ./ gc )
       .* ( -((d - dxyes * bbd) / sigma)
        + rho * ( - sxyes * bbs - rho * (d - dxyes * bbd)
        / sigma) / (1 - rho^2) )  );
    hb1 = sumc( - hp ./ hc  );
    hb2 = sumc(  (hp ./ hc) .* ( - sxnom[.,2]
        + sxnom * bbs .* (2 * b2 * sxnov[.,1]
        + 2 * b3 * sxnov[.,3]) ./ (2 * h ) )  );
    hb3 = sumc(  (hp ./ hc) .* ( - sxnom[.,3]
        + sxnom * bbs .* (2 * b3 * sxnov[.,2]
        + 2 * b2 * sxnov[.,3]) ./ (2 * h ) )  );

    k = (gbbs'+(hb1~hb2~hb3)) ~ (fbbd + gbbd)'
      ~ (fss + gss) ~ grho;

    retp(-k);
endp;


iter = iter + 1;
endo;

system;
```

APPENDIX E

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A TOBIT DEMAND EQUATION

APPENDIX E

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A TOBIT DEMAND EQUATION

This GAUSS program is used to conduct Monte Carlo experiments for the self-selection model with measurement errors and a Tobit demand equation (ρ = 0.25).

```
new;
use optmum;
output file = tt.out;

@-- declare global variables --@
declare matrix g_bbd, g_bbs, g_sigma, g_rho;
external matrix g_bbd, g_bbs, g_sigma, g_rho;

iter = 1;
do while iter le 500;

recal:

@--
npop: population size
nx: number of variables (x's and e's)
nb: number of blocks (nb)
bobs: number of obs in a block
ss: number of obs used to calculate variance-covariance
smp: random sub-sample / population (0 < = smp < = 1)
--@
npop = 10000;
nx = 5;
nb = 250;
bobs = npop / nb;
ss = 200;
smp = 0.1;
```

```
@-- draw random sample as population --@
x = rndn(npop,nx);
let v[5,5] = 1.44   0.24  0.096  0    0
             0.24   1     0.24   0    0
             0.096  0.24  0.64   0    0
             0      0     0      1    0.25
             0      0     0      0.25 1;
sqrtv = chol(v);
x = x * sqrtv;
let a[1,5] = 3 1.5 4 0 0;
x = x + a;
evar = vcx(x[1:ss,1 2]);
clear v, sqrtv, a;


@-- divide population into blocks --@
si = 2000;
s = npop / si;
sg = nb / s;
w = eye(sg) .*. ones(bobs,1);
sx = x[1:si,1 2];
msx = (sx'w/bobs)' .*. ones(bobs,1);
vsx = ((((sx-msx)^2)
      ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
     w / (bobs-1) )' .*. ones(bobs,1);
xx = msx~vsx;
i = 2;
do while i <= s;
    l = si * (i - 1) + 1;
    h = si * i;
    sx = x[l:h,1 2];
    msx = (sx'w/bobs)' .*. ones(bobs,1);
    vsx = ((((sx-msx)^2)
          ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
         w / (bobs-1) )' .*. ones(bobs,1);
    sx = msx~vsx;
    xx = xx|sx;
    i = i + 1;
endo;
clear i, l, h, msx, vsx, sx;


@-- xx = [x1 x2 x3 e1 e2 m1 m2 v11 v22 v21] --@
x = x~xx;
clear xx;


@--
extract a random sub-sample from population xx,
and variables in the random sub-sample are
[x1 x2 x3 e1 e2 m1 m2 v11 v22 v21].
--@
```

```
i = rndu(npop,1);
x = i~x;
data = selif(x,x[.,1] .le smp);
clear x;
data = data[.,2:(cols(data))];
obs = rows(data);


@--
generate dependent variables S* and D*
S* = 1.5 + 1 * x1 - 3 * x2 + e1
D* = 6 + 4 * x2 - 3 * x3 + e2
--@
let bs[3,1] = 1.5  1  -3;
let bd[3,1] = 6  4  -3;
s = (ones(obs,1)~data[.,1 2]) * bs + data[.,4];
d = (ones(obs,1)~data[.,2 3]) * bd + data[.,5];


@--
rearrange data according to
s > 0: xyes = [d x1 x2 x3], and
s <= 0: xno = [m1 m2 v11 v22 v21]
also data = [s d x1 x2 x3 m1 m2 v11 v22 v21].
--@
data = s~d~data[.,1:3 6:10];
clear s, d;

s1 = (data[.,1] .gt 0) .and (data[.,2] .gt 0);
s0 = (data[.,1] .gt 0) .and (data[.,2] .le 0);
yes1 = selif(data,s1);
yes0 = selif(data,s0);
no = selif(data, data[.,1] .le 0);
clear s1,s0,data;

nno = rows(no);
nyes1 = rows(yes1);
nyes0 = rows(yes0);
nyes = nyes1 + nyes0;

sxyes1 = ones(nyes1,1)~yes1[.,3 4];
dxyes1 = ones(nyes1,1)~yes1[.,4 5];
d = yes1[.,2];
sxyes0 = ones(nyes0,1)~yes0[.,3 4];
dxyes0 = ones(nyes0,1)~yes0[.,4 5];
clear yes1, yes0;

sxno = ones(nno,1)~no[.,3 4];
sxnom = ones(nno,1)~no[.,6 7];
sxnov = no[.,8 9 10];
clear no;
```

```
@--
starting values for
[s0 s1 s2 d0 d1 d2 sigma^2 rho]
--@
bt = bd|1;
b0 = bs|bd|1|0.25;

@-- call and print optmum --@

optset;
 opgdprc = &foc1;
{beta1, f1, g, retcode} = optmum(&fn1, b0);
if retcode ne 0;
    goto recal;
endif;
cov1 = invpd(hessp(&fn1,beta1));
stder1 = sqrt(diag(cov1));
pout = (-f1)~beta1'~stder1';

optset;
 opgdprc = &foc3;
{beta3, f3, g, retcode} = optmum(&fn3, beta1);
if retcode ne 0;
    goto recal;
endif;
cov3 = invpd(hessp(&fn3,beta3));
stder3 = sqrt(diag(cov3));
pout = pout~(-f3)~beta3'~stder3';

optset;
 opgdprc = &foc4;
{beta4, f4, g, retcode} = optmum(&fn4, beta3);
if retcode ne 0;
    goto recal;
endif;
cov4 = invpd(hessp(&fn4,beta4));
stder4 = sqrt(diag(cov4));
pout = pout~(-f4)~beta4'~stder4';

optset;
 opgdprc = &foc0;
{beta0, f0, g, retcode} = optmum(&fn0, bt);
if retcode ne 0;
    goto recal;
endif;
cov0 = invpd(hessp(&fn0,beta0));
stder0 = sqrt(diag(cov0));
output on;
iter~nyes0~nyes1~nyes~nno~(-f0)~beta0'~stder0'~pout;
print " ";
output off;
```

```
@-- procedures --@
@--
-(log-likelihood) function for a Tobit demand function
using data from only the respondents
--@
proc fn0(para);
    local bbd, sigma, l1, l2;
    bbd = para[1 2 3,.];
    sigma = sqrt(para[4,.]);
    l1 = (1/sigma) * pdfn((d - dxyes1 * bbd) / sigma);
    l2 = cdfn( (- dxyes0 * bbd) / sigma );
    retp( - sumc( ln(l1|l2) ) );
endp;

proc foc0(para);
    local bbd, s, fb, fs2, f;
    bbd = para[1 2 3,.];
    s = sqrt(para[4,.]);
    fb = sumc( (pdfn( (-dxyes0*bbd) / s)
            ./ cdfn( (-dxyes0*bbd) / s) )
            .* (dxyes0/s) )'
            - (1/s^2) * ( d-dxyes1*bbd )' dxyes1;
    fs2 = sumc( 0.5 * (pdfn( (-dxyes0*bbd) / s )
            ./ cdfn( (-dxyes0*bbd) / s) )
            .* ( (-dxyes0*bbd) / (s^3) ) )'
            + (nyes1 / (2*s^2) )
            - (1 / (2*s^4) ) * (d-dxyes1*bbd)'(d-dxyes1*bbd);
    f = fb~fs2;
    retp(f);
endp;


@--
-(log-likelihood) function for a self-selection model
with a Tobit demand function using censored data
(non-respondents' independent variables are observed)
--@
proc fn1(para);
    local bbs, bbd, sigma, rho, no, yes0, yes1;
    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    no = cdfn( - sxno * bbs );
    yes0 = cdfn( (- dxyes0 * bbd) / sigma ) -
        cdfbvn( ((- dxyes0 * bbd) / sigma),
            (- sxyes0 * bbs), rho);
    yes1 = (1/sigma) * pdfn((d - dxyes1 * bbd) / sigma)
        .* cdfnc( (- sxyes1 * bbs -
            (rho / sigma) * (d - dxyes1 * bbd) )
            / sqrt( 1 - rho^2) );
    retp(-sumc( ln(no|yes0|yes1) ));
endp;
```

```
proc foc1(para);
    local bbs, bbd, sigma, rho, f, cyes1,
        gp, gc, cno, hp, hc, fbbd, fss,
        gbbd, gss, gbbs, grho, hbbs, k,
        ip, ic, bc, ibc, iyes0a, iyes0b,
        ibbd, iss, ibbs, bvd, irho;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    f = (1 / sigma) * pdfn( (d - dxyes1 * bbd) / sigma);
    cyes1 = ( - sxyes1 * bbs
            - rho * (d - dxyes1 *bbd) / sigma )
          / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes1);
    gc = cdfnc(cyes1);
    hp = pdfn( - sxno * bbs);
    hc = cdfn( - sxno * bbs);

    fbbd = sumc(
          ((d - dxyes1 * bbd) / sigma^2) .* dxyes1 );
    fss = sumc(
         ((d - dxyes1 * bbd)^2 - sigma^2)
          / (2 * sigma^4)    );
    gbbd = sumc(
         (- gp ./ gc) * (rho / sigma) .* dxyes1 );
    gss = sumc(
         (- gp ./ gc) .* (rho * (d - dxyes1 * bbd))
         / (2 * sigma^3)   );
    gbbs = sumc(  ( gp ./ gc) .*  sxyes1   );
    grho = sumc(   (- gp ./ gc )
        .* ( -((d - dxyes1 * bbd) / sigma)
         + rho * ( - sxyes1 * bbs - rho * (d - dxyes1 * bbd)
         / sigma) / (1 - rho^2) )  );
    hbbs = sumc( (hp ./ hc) .* (- sxno) );

    g_bbs = bbs;
    g_bbd = bbd;
    g_sigma = sigma;
    g_rho = rho;
    ip = pdfn(- dxyes0 * bbd / sigma);
    ic = cdfn(- dxyes0 * bbd / sigma);
    ibc =  ic -
        cdfbvn( ((- dxyes0 * bbd) / sigma),
              (- sxyes0 * bbs), rho);
    iyes0a = ( - sxyes0 * bbs
            - rho * (- dxyes0 * bbd) / sigma)
          / sqrt(1 - rho^2);
    iyes0b = ( - dxyes0 * bbd
            - rho * sigma * (- sxyes0 * bbs) )
          / sqrt(1 - rho^2);
    ibbd = sumc( (ip .* (- dxyes0 / sigma)
```

```
            - (1 / sigma) * ip .* cdfn(iyes0a)
            .* (- dxyes0 / sigma) ) ./ ibc);
      iss = sumc( (ip * .5 .* (dxyes0 * bbd / sigma^(3/2) )
            - (1 / sigma) * ip .* cdfn(iyes0a) * (-.5)
            .* (- dxyes0 * bbd / sigma^(3/2) ) )
            ./ ibc);
      ibbs = sumc( - pdfn(- sxyes0 * bbs) .* cdfn(iyes0b)
            .* (- sxyes0) ./ ibc);
      bvd = gradp(&bvn,g_rho);
      irho = sumc( - bvd ./ ibc);

      k = (gbbs' + hbbs' + ibbs') ~ (fbbd + gbbd + ibbd)'
          ~ (fss + gss + iss) ~ (grho + irho);

      retp(-k);
   endp;


   @--
   -(log-likelihood) function for a self-selection model
   with measurement errors and a Tobit demand function
   using empirical block mean and estimateted variance-
   covariance (200 obs. from the population)
   for measurement error
   --@
   proc fn3(para);
      local bbs, bbd, sigma, rho, bbsn, delta,
            no, yes0, yes1;
      bbs = para[1 2 3,.];
      bbd = para[4 5 6,.];
      sigma = sqrt(para[7,.]);
      rho = para[8,.];
      bbsn = bbs[2 3,.];
      delta = sqrt(1 + bbsn'evar*bbsn);
      no = cdfn( - sxnom * bbs / delta );
      yes0 = cdfn( (- dxyes0 * bbd) / sigma ) -
            cdfbvn( ((- dxyes0 * bbd) / sigma),
                   (- sxyes0 * bbs), rho);
      yes1 = (1/sigma) * pdfn((d - dxyes1 * bbd) / sigma)
            .* cdfnc( (- sxyes1 * bbs -
                   (rho / sigma) * (d - dxyes1 * bbd) )
                   / sqrt( 1 - rho^2) );
      retp(-sumc( ln(no|yes0|yes1) ));
   endp;

   proc foc3(para);
      local bbs, b1, b2, b3, bbd, sigma, rho, f, cyes1,
            gp, gc, cno, hp, hc, h, fbbd, fss,
            gbbd, gss, gbbs, grho, hb1, hb2, hb3, k,
            ip, ic, bc, ibc, iyes0a, iyes0b,
            ibbd, iss, ibbs, bvd, irho;

      bbs = para[1 2 3,.];
```

```
    b1 = bbs[1,.];
    b2 = bbs[2,.];
    b3 = bbs[3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    f = (1 / sigma) * pdfn( (d - dxyes1 * bbd) / sigma);
    cyes1 = ( - sxyes1 * bbs
            - rho * (d - dxyes1 *bbd) / sigma )
          / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes1);
    gc = cdfnc(cyes1);
    cno = (- sxnom * bbs) ./
          sqrt(1 + b2^2 * evar[1,1] + b3^2 * evar[2,2]
               + 2 * b2 * b3 * evar[1,2]);
    hp = (1 / sqrt(1 + b2^2 * evar[1,1]
                 + b3^2 * evar[2,2]
                 + 2 * b2 * b3 * evar[1,2]))
         .* pdfn(cno);
    hc = cdfn(cno);
    h =  1 + b2^2 * evar[1,1]
       + b3^2 * evar[2,2] + 2 * b2 * b3 * evar[1,2];

    fbbd = sumc(
          ((d - dxyes1 * bbd) / sigma^2) .* dxyes1  );
    fss = sumc(
         ((d - dxyes1 * bbd)^2 - sigma^2)
          / (2 * sigma^4)    );
    gbbd = sumc(
         (- gp ./ gc) * (rho / sigma) .* dxyes1  );
    gss = sumc(
         (- gp ./ gc) .* (rho * (d - dxyes1 * bbd))
         / (2 * sigma^3)   );
    gbbs = sumc(  ( gp ./ gc) .*  sxyes1   );
    grho = sumc(  (- gp ./ gc )
        .* ( -((d - dxyes1 * bbd) / sigma)
         + rho * ( - sxyes1 * bbs - rho * (d - dxyes1 * bbd)
         / sigma) / (1 - rho^2) )  );
    hb1 = sumc( - hp ./ hc  );
    hb2 = sumc(  (hp ./ hc) .* ( - sxnom[.,2]
         + sxnom * bbs .* (2 * b2 * evar[1,1]
         + 2 * b3 * evar[1,2]) ./ (2 * h ) )  );
    hb3 = sumc(  (hp ./ hc) .* ( - sxnom[.,3]
         + sxnom * bbs .* (2 * b3 * evar[2,2]
         + 2 * b2 * evar[1,2]) ./ (2 * h ) )  );

    g_bbs = bbs;
    g_bbd = bbd;
    g_sigma = sigma;
    g_rho = rho;
    ip = pdfn(- dxyes0 * bbd / sigma);
    ic = cdfn(- dxyes0 * bbd / sigma);
    ibc =  ic -
```

```
        cdfbvn( ((- dxyes0 * bbd) / sigma),
                (- sxyes0 * bbs), rho);
    iyes0a = ( - sxyes0 * bbs
             - rho * (- dxyes0 * bbd) / sigma)
           / sqrt(1 - rho^2);
    iyes0b = ( - dxyes0 * bbd
             - rho * sigma * (- sxyes0 * bbs) )
           / sqrt(1 - rho^2);
    ibbd = sumc( (ip .* (- dxyes0 / sigma)
         - (1 / sigma) * ip .* cdfn(iyes0a)
         .* (- dxyes0 / sigma) ) ./ ibc);
    iss = sumc( (ip * .5 .* (dxyes0 * bbd / sigma^(3/2) )
        - (1 / sigma) * ip .* cdfn(iyes0a) * (-.5)
        .* (- dxyes0 * bbd / sigma^(3/2) ) )
        ./ ibc);
    ibbs = sumc( - pdfn(- sxyes0 * bbs) .* cdfn(iyes0b)
         .* (- sxyes0) ./ ibc);
    bvd = gradp(&bvn,g_rho);
    irho = sumc( - bvd ./ ibc);

    k = (gbbs'+(hb1~hb2~hb3)+ibbs') ~ (fbbd + gbbd + ibbd)'
        ~ (fss + gss + iss) ~ (grho + irho);

    retp(-k);
endp;


@--
-(log-likelihood) function for a self-selection model
with measurement errors and a Tobit demand function
using empirical block mean and empirical block
variance-covariance for measurement error
--@
proc fn4(para);
    local bbs, bbd, sigma, rho, bbs1, bbs2,
        delta, no, yes0, yes1;
    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    bbs1 = bbs[2,.];
    bbs2 = bbs[3,.];
    delta = sqrt(1 + bbs1^2 * sxnov[.,1]
                + bbs2^2 * sxnov[.,2]
                + 2 * bbs1 * bbs2 * sxnov[.,3]);
    no = cdfn( - sxnom * bbs ./ delta );
    yes0 = cdfn( (- dxyes0 * bbd) / sigma ) -
        cdfbvn( ((- dxyes0 * bbd) / sigma),
                (- sxyes0 * bbs), rho);
    yes1 = (1/sigma) * pdfn((d - dxyes1 * bbd) / sigma)
        .* cdfnc( (- sxyes1 * bbs -
                (rho / sigma) * (d - dxyes1 * bbd) )
                / sqrt( 1 - rho^2) );
```

```
    retp(-sumc( ln(no|yes0|yes1) ));
endp;

proc foc4(para);
    local bbs, b1, b2, b3, bbd, sigma, rho, f, cyes1,
          gp, gc, cno, hp, hc, h, fbbd, fss,
          gbbd, gss, gbbs, grho, hb1, hb2, hb3, k,
          ip, ic, bc, ibc, iyes0a, iyes0b,
          ibbd, iss, ibbs, bvd, irho;

    bbs = para[1 2 3,.];
    b1 = bbs[1,.];
    b2 = bbs[2,.];
    b3 = bbs[3,.];
    bbd = para[4 5 6,.];
    sigma = sqrt(para[7,.]);
    rho = para[8,.];
    f = (1 / sigma) * pdfn( (d - dxyes1 * bbd) / sigma);
    cyes1 = ( - sxyes1 * bbs
            - rho * (d - dxyes1 *bbd) / sigma )
          / sqrt(1 - rho^2);
    gp = (1 / sqrt(1 - rho^2)) * pdfn(cyes1);
    gc = cdfnc(cyes1);
    cno = (- sxnom * bbs) ./
         sqrt(1 + b2^2 * sxnov[.,1] + b3^2 * sxnov[.,2]
              + 2 * b2 * b3 * sxnov[.,3]);
    hp = (1 / sqrt(1 + b2^2 * sxnov[.,1]
                 + b3^2 * sxnov[.,2]
                 + 2 * b2 * b3 * sxnov[.,3]))
       .* pdfn(cno);
    hc = cdfn(cno);
    h =  1 + b2^2 * sxnov[.,1]
       + b3^2 * sxnov[.,2] + 2 * b2 * b3 * sxnov[.,3];

    fbbd = sumc(
         ((d - dxyes1 * bbd) / sigma^2) .* dxyes1  );
    fss = sumc(
         ((d - dxyes1 * bbd)^2 - sigma^2)
          / (2 * sigma^4)    );
    gbbd = sumc(
         (- gp ./ gc) * (rho / sigma) .* dxyes1  );
    gss = sumc(
         (- gp ./ gc) .* (rho * (d - dxyes1 * bbd))
         / (2 * sigma^3)   );
    gbbs = sumc(   ( gp ./ gc) .*  sxyes1   );
    grho = sumc(   (- gp ./ gc )
       .* ( -((d - dxyes1 * bbd) / sigma)
        + rho * ( - sxyes1 * bbs - rho * (d - dxyes1 * bbd)
        / sigma) / (1 - rho^2) )  );
    hb1 = sumc( - hp ./ hc  );
    hb2 = sumc(  (hp ./ hc) .* ( - sxnom[.,2]
        + sxnom * bbs .* (2 * b2 * sxnov[.,1]
        + 2 * b3 * sxnov[.,3]) ./ (2 * h ) )  );
```

```
        hb3 = sumc( (hp ./ hc) .* ( - sxnom[.,3]
              + sxnom * bbs .* (2 * b3 * sxnov[.,2]
              + 2 * b2 * sxnov[.,3]) ./ (2 * h ) ) );

        g_bbs = bbs;
        g_bbd = bbd;
        g_sigma = sigma;
        g_rho = rho;
        ip = pdfn(- dxyes0 * bbd / sigma);
        ic = cdfn(- dxyes0 * bbd / sigma);
        ibc = ic -
              cdfbvn( ((- dxyes0 * bbd) / sigma),
                     (- sxyes0 * bbs), rho);
        iyes0a = ( - sxyes0 * bbs
                 - rho * (- dxyes0 * bbd) / sigma)
               / sqrt(1 - rho^2);
        iyes0b = ( - dxyes0 * bbd
                 - rho * sigma * (- sxyes0 * bbs) )
               / sqrt(1 - rho^2);
        ibbd = sumc( (ip .* (- dxyes0 / sigma)
               - (1 / sigma) * ip .* cdfn(iyes0a)
               .* (- dxyes0 / sigma) ) ./ ibc);
        iss = sumc( (ip * .5 .* (dxyes0 * bbd / sigma^(3/2) )
              - (1 / sigma) * ip .* cdfn(iyes0a) * (-.5)
              .* (- dxyes0 * bbd / sigma^(3/2) ) )
              ./ ibc);
        ibbs = sumc( - pdfn(- sxyes0 * bbs) .* cdfn(iyes0b)
               .* (- sxyes0) ./ ibc);
        bvd = gradp(&bvn,g_rho);
        irho = sumc( - bvd ./ ibc);

        k = (gbbs'+(hb1~hb2~hb3)+ibbs') ~ (fbbd + gbbd + ibbd)'
            ~ (fss + gss + iss) ~ (grho + irho);

        retp(-k);
    endp;


    proc bvn(r);
        local r0;
        r0 = r;
        retp( cdfbvn( ((- dxyes0 * g_bbd) / g_sigma),
                      (- sxyes0 * g_bbs), r0) );
    endp;

    iter = iter + 1;
    endo;

    system;
```

APPENDIX F

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A PROBIT DEMAND EQUATION

APPENDIX F

A GAUSS PROGRAM FOR MONTE CARLO EXPERIMENTS: SELF-SELECTION MODEL WITH MEASUREMENT ERRORS AND A PROBIT DEMAND EQUATION

This GAUSS program is used to conduct Monte Carlo experiments for the self-selection model with measurement errors and a probit demand equation (ρ = 0.25).

```
new;
use optmum;
output file = tp.out;

@-- declare global variables --@
declare matrix g_bbd, g_bbs, g_rho;
external matrix g_bbd, g_bbs, g_rho;

iter = 1;
do while iter le 500;

recal:

@--
npop: population size
nx: number of variables (x's and e's)
nb: number of blocks (nb)
bobs: number of obs in a block
ss: number of obs used to calculate variance-covariance
smp: random sub-sample / population (0 <= smp <= 1)
--@
npop = 10000;
nx = 5;
nb = 250;
bobs = npop / nb;
ss = 200;
smp = 0.1;
```

```
@-- draw random sample as population --@
x = rndn(npop,nx);
let v[5,5] = 1.44   0.24  0.096  0   0
             0.24   1     0.24   0   0
             0.096  0.24  0.64   0   0
             0      0     0      1   0.25
             0      0     0      0.25  1;
sqrtv = chol(v);
x = x * sqrtv;
let a[1,5] = 3 1.5 4 0 0;
x = x + a;
evar = vcx(x[1:ss,1 2]);
clear v, sqrtv, a;


@-- divide population into blocks --@
si = 2000;
s = npop / si;
sg = nb / s;
w = eye(sg) .*. ones(bobs,1);
sx = x[1:si,1 2];
msx = (sx'w/bobs)' .*. ones(bobs,1);
vsx = ((((sx-msx)^2)
      ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
     w / (bobs-1) )' .*. ones(bobs,1);
xx = msx~vsx;
i = 2;
do while i <= s;
    l = si * (i - 1) + 1;
    h = si * i;
    sx = x[l:h,1 2];
    msx = (sx'w/bobs)' .*. ones(bobs,1);
    vsx = ((((sx-msx)^2)
          ~ ((sx[.,1]-msx[.,1]).*(sx[.,2]-msx[.,2])))'
         w / (bobs-1) )' .*. ones(bobs,1);
    sx = msx~vsx;
    xx = xx|sx;
    i = i + 1;
endo;
clear i, l, h, msx, vsx, sx;


@-- xx = [x1 x2 x3 e1 e2 m1 m2 v11 v22 v21] --@
x = x~xx;
clear xx;


@--
extract a random sub-sample from population xx,
and variables in the random sub-sample are
[x1 x2 x3 e1 e2 m1 m2 v11 v22 v21].
--@
```

```
i = rndu(npop,1);
x = i~x;
data = selif(x,x[.,1] .le smp);
clear x;
data = data[.,2:(cols(data))];
obs = rows(data);


@--
generate dependent variables S* and D*
S* = 1.5 + 1 * x1 - 3 * x2 + e1
D* = 6 + 4 * x2 - 3 * x3 + e2
--@
let bs[3,1] = 1.5  1  -3;
let bd[3,1] = 6  4  -3;
s = (ones(obs,1)~data[.,1 2]) * bs + data[.,4];
d = (ones(obs,1)~data[.,2 3]) * bd + data[.,5];


@--
rearrange data according to
s > 0: xyes = [d x1 x2 x3], and
s <= 0: xno = [m1 m2 v11 v22 v21]
also data = [s d x1 x2 x3 m1 m2 v11 v22 v21].
--@
data = s~d~data[.,1:3 6:10];
clear s, d;

s1 = (data[.,1] .gt 0) .and (data[.,2] .gt 0);
s0 = (data[.,1] .gt 0) .and (data[.,2] .le 0);
yes1 = selif(data,s1);
yes0 = selif(data,s0);
no = selif(data, data[.,1] .le 0);
clear s1,s0,data;

nno = rows(no);
nyes1 = rows(yes1);
nyes0 = rows(yes0);
nyes = nyes1 + nyes0;

sxyes1 = ones(nyes1,1)~yes1[.,3 4];
dxyes1 = ones(nyes1,1)~yes1[.,4 5];
sxyes0 = ones(nyes0,1)~yes0[.,3 4];
dxyes0 = ones(nyes0,1)~yes0[.,4 5];
clear yes1, yes0;

sxno = ones(nno,1)~no[.,3 4];
sxnom = ones(nno,1)~no[.,6 7];
sxnov = no[.,8 9 10];
clear no;


@--
```

```
starting values for
[s0 s1 s2 d0 d1 d2 sigma^2 rho]
--@
bp = bd;
b0 = bs|bd|0.25;


@-- call and print optmum --@

optset;
 opgdprc = &foc1;
{beta1, f1, g, retcode} = optmum(&fn1, b0);
if retcode ne 0;
    goto recal;
endif;
cov1 = invpd(hessp(&fn1,beta1));
stder1 = sqrt(diag(cov1));
pout = (-f1)~beta1'~stder1';

optset;
 opgdprc = &foc3;
{beta3, f3, g, retcode} = optmum(&fn3, beta1);
if retcode ne 0;
    goto recal;
endif;
cov3 = invpd(hessp(&fn3,beta3));
stder3 = sqrt(diag(cov3));
pout = pout~(-f3)~beta3'~stder3';

optset;
 opgdprc = &foc4;
{beta4, f4, g, retcode} = optmum(&fn4, beta3);
if retcode ne 0;
    goto recal;
endif;
cov4 = invpd(hessp(&fn4,beta4));
stder4 = sqrt(diag(cov4));
pout = pout~(-f4)~beta4'~stder4';

optset;
 opgdprc = &foc0;
{beta0, f0, g, retcode} = optmum(&fn0, bp);
if retcode ne 0;
    goto recal;
endif;
cov0 = invpd(hessp(&fn0,beta0));
stder0 = sqrt(diag(cov0));
output on;
iter~nyes0~nyes1~nyes~nno~(-f0)~beta0'~stder0'~pout;
print " ";
output off;
```

```
@-- procedures --@

@--
-(log-likelihood) function for a probit demand function
using data from only the respondents
--@
proc fn0(para);
     local bb, l1, l2;
     bb = para;
     l1 = cdfn(dxyes1 * bb);
     l2 = cdfn(- dxyes0 * bb);
     retp( - sumc( ln(l1|l2) ) );
endp;

proc foc0(para);
     local bb, y, z, ff;
     bb = para;
     y = ones(nyes1,1)|zeros(nyes0,1);
     x = dxyes1|dxyes0;
     ff = sumc(
          ( (y - cdfn(x * bb)) .* pdfn(x * bb) .* x)
          ./ (cdfn(x * bb) .* cdfnc(x * bb)) )';
     retp(-ff);
endp;


@--
-(log-likelihood) function for a self-selection model
with a probit demand function using censored data
(non-respondents' independent variables are observed)
--@
proc fn1(para);
     local bbs, bbd, rho, no, yes0, yes1, ll;
     bbs = para[1 2 3,.];
     bbd = para[4 5 6,.];
     rho = para[7,.];
     no = cdfn( - sxno * bbs );
     yes0 = cdfn( - dxyes0 * bbd) -
          cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
     yes1 = cdfnc( (- sxyes1 * bbs) ) -
          ( cdfn( (- dxyes1 * bbd) ) -
            cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
     trap 1;
     ll = ln(no|yes0|yes1);
     if scalerr(ll);
        ll = "NAN";
     endif;
     retp(-sumc( ll ));
endp;

proc foc1(para);
     local bbs, bbd, rho, fbbs,
           den0, gbbd, gbbs,
```

```
            den1, hbbd, hbbs,
            bvr0, bvr1, grho, hrho, k;

    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    rho = para[7,.];

    fbbs= sumc( (pdfn(- sxno * bbs)
            ./ cdfn(- sxno * bbs)) .* (-sxno) );

    den0 = cdfn( - dxyes0 * bbd) -
          cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
    gbbd = sumc( (1/den0) .*
             cdfnc( ( -sxyes0*bbs - (rho * (-dxyes0*bbd)))
                  / sqrt(1 - rho^2) ) .*
             pdfn( -dxyes0*bbd ) .* (-dxyes0) );
    gbbs = sumc( (1/den0) .* (- pdfn( -sxyes0*bbs ) .*
             cdfn( ( -dxyes0*bbd - (rho * (-sxyes0*bbs)))
                  / sqrt(1 - rho^2) ) .*
             (-sxyes0) ) );
    den1 = cdfnc(- sxyes1 * bbs) - ( cdfn(- dxyes1 * bbd) -
          cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
    hbbd = sumc( (-1/den1) .*
             cdfnc( ( -sxyes1*bbs - (rho * (-dxyes1*bbd)))
                  / sqrt(1 - rho^2) ) .*
             pdfn( -dxyes1*bbd ) .* (-dxyes1) );
    hbbs = sumc( (1/den1) .*
             ( - pdfn(- sxyes1*bbs) + pdfn( -sxyes1*bbs ) .*
             cdfn( ( -dxyes1*bbd - (rho * (-sxyes1*bbs)))
                  / sqrt(1 - rho^2) ) ) .* (-sxyes1) );
    g_bbs = bbs;
    g_bbd = bbd;
    g_rho = rho;
    bvr0 = gradp(&bvn0,g_rho);
    bvr1 = gradp(&bvn1,g_rho);
    grho = sumc( (1/den0) .* (- bvr0) );
    hrho = sumc( (1/den1) .* bvr1 );

    k = (fbbs' + gbbs' + hbbs') ~ (gbbd + hbbd)'
        ~ (grho + hrho);

    retp(-k);
endp;


@--
-(log-likelihood) function for a self-selection model
with measurement errors and a probit demand function
using empirical block mean and estimateted variance-
covariance (200 obs. from the population)
for measurement error
--@
proc fn3(para);
```

```
    local bbs, bbd, rho, bbsn, delta,
          no, yes0, yes1, ll;
    bbs = para[1 2 3,.];
    bbd = para[4 5 6,.];
    rho = para[7,.];
    bbsn = bbs[2 3,.];
    delta = sqrt(1 + bbsn'evar*bbsn);
    no = cdfn( - sxnom * bbs / delta );
    yes0 = cdfn( - dxyes0 * bbd) -
          cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
    yes1 = cdfnc( (- sxyes1 * bbs) ) -
          ( cdfn( (- dxyes1 * bbd) ) -
            cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
    ll = ln(no|yes0|yes1);
    retp(-sumc( ll ));
endp;

proc foc3(para);
    local bbs, b1, b2, b3, bbd, rho,
          delta, z,
          fb1, fb2, fb3,
          den0, gbbd, gbbs,
          den1, hbbd, hbbs,
          bvr0, bvr1, grho, hrho, k;

    bbs = para[1 2 3,.];
    b1 = bbs[1,.];
    b2 = bbs[2,.];
    b3 = bbs[3,.];
    bbd = para[4 5 6,.];
    rho = para[7,.];
    delta = sqrt(1 + b2^2 * evar[1,1] + b3^2 * evar[2,2]
                 + 2 * b2 * b3 * evar[1,2]);
    z = -sxnom*bbs / delta;
    fb1 = sumc( (- pdfn(z) ./ cdfn(z) ) / delta );
    fb2 = sumc( ( pdfn(z) ./ cdfn(z) ) .* ( - sxnom[.,2] / delta
          + sxnom * bbs .* (2 * b2 * evar[1,1]
          + 2 * b3 * evar[1,2]) ./ (2 * delta^3 ) ) );
    fb3 = sumc( ( pdfn(z) ./ cdfn(z) ) .* ( - sxnom[.,3] / delta
          + sxnom * bbs .* (2 * b3 * evar[2,2]
          + 2 * b2 * evar[1,2]) ./ (2 * delta^3 ) ) );
    den0 = cdfn( - dxyes0 * bbd) -
          cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
    gbbd = sumc( (1/den0) .*
              cdfnc( ( -sxyes0*bbs - (rho * (-dxyes0*bbd)))
                    / sqrt(1 - rho^2) ) .*
              pdfn( -dxyes0*bbd ) .* (-dxyes0) );
    gbbs = sumc( (1/den0) .* (- pdfn( -sxyes0*bbs ) .*
              cdfn( ( -dxyes0*bbd - (rho * (-sxyes0*bbs)))
                    / sqrt(1 - rho^2) ) .*
             (-sxyes0) ) );
    den1 = cdfnc(- sxyes1 * bbs) - ( cdfn(- dxyes1 * bbd) -
          cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
```

```
        hbbd = sumc( (-1/den1) .*
                  cdfnc( ( -sxyes1*bbs - (rho * (-dxyes1*bbd)))
                       / sqrt(1 - rho^2) ) .*
                  pdfn( -dxyes1*bbd ) .* (-dxyes1) );
        hbbs = sumc( (1/den1) .*
                  ( - pdfn(- sxyes1*bbs) + pdfn( -sxyes1*bbs ) .*
                   cdfn( ( -dxyes1*bbd - (rho * (-sxyes1*bbs)))
                       / sqrt(1 - rho^2) ) ) .* (-sxyes1) );
        g_bbs = bbs;
        g_bbd = bbd;
        g_rho = rho;
        bvr0 = gradp(&bvn0,g_rho);
        bvr1 = gradp(&bvn1,g_rho);
        grho = sumc( (1/den0) .* (- bvr0) );
        hrho = sumc( (1/den1) .* bvr1 );

        k = ((fb1~fb2~fb3) + gbbs'+ hbbs')
           ~ (gbbd + hbbd)' ~ (grho + hrho);

        retp(-k);
    endp;


    @--
    -(log-likelihood) function for a self-selection model
    with measurement errors and a probit demand function
    using empirical block mean and empirical block
    variance-covariance for measurement error
    --@
    proc fn4(para);
        local bbs, bbd, rho, bbs1, bbs2,
              delta, no, yes0, yes1, ll;
        bbs = para[1 2 3,.];
        bbd = para[4 5 6,.];
        rho = para[7,.];
        bbs1 = bbs[2,.];
        bbs2 = bbs[3,.];
        delta = sqrt(1 + bbs1^2 * sxnov[.,1]
                     + bbs2^2 * sxnov[.,2]
                     + 2 * bbs1 * bbs2 * sxnov[.,3]);
        no = cdfn( - sxnom * bbs ./ delta );
        yes0 = cdfn( - dxyes0 * bbd) -
              cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
        yes1 = cdfnc( (- sxyes1 * bbs) ) -
              ( cdfn( (- dxyes1 * bbd) ) -
                cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
        ll = ln(no|yes0|yes1);
        retp(-sumc( ll ));
    endp;

    proc foc4(para);
        local bbs, b1, b2, b3, bbd, rho,
              delta, z,
```

```
        fb1, fb2, fb3,
        den0, gbbd, gbbs,
        den1, hbbd, hbbs,
        bvr0, bvr1, grho, hrho, k;

  bbs = para[1 2 3,.];
  b1 = bbs[1,.];
  b2 = bbs[2,.];
  b3 = bbs[3,.];
  bbd = para[4 5 6,.];
  rho = para[7,.];
  delta = sqrt(1 + b2^2 * sxnov[.,1] + b3^2 * sxnov[.,2]
           + 2 * b2 * b3 * sxnov[.,3]);
  z = -sxnom*bbs ./ delta;
  fb1 = sumc( (- pdfn(z) ./ cdfn(z) ) ./ delta );
  fb2 = sumc( ( pdfn(z) ./ cdfn(z) ) .* ( - sxnom[.,2] ./ delta
       + sxnom * bbs .* (2 * b2 * sxnov[.,1]
       + 2 * b3 * sxnov[.,3]) ./ (2 * delta^3 ) )  );
  fb3 = sumc( ( pdfn(z) ./ cdfn(z) ) .* ( - sxnom[.,3] ./ delta
       + sxnom * bbs .* (2 * b3 * sxnov[.,2]
       + 2 * b2 * sxnov[.,3]) ./ (2 * delta^3 ) )  );
  den0 = cdfn( - dxyes0 * bbd) -
       cdfbvn( (- dxyes0 * bbd), (- sxyes0 * bbs), rho);
  gbbd = sumc( (1/den0) .*
           cdfnc( ( -sxyes0*bbs - (rho * (-dxyes0*bbd)))
                / sqrt(1 - rho^2) ) .*
           pdfn( -dxyes0*bbd ) .* (-dxyes0) );
  gbbs = sumc( (1/den0) .* (- pdfn( -sxyes0*bbs ) .*
           cdfn( ( -dxyes0*bbd - (rho * (-sxyes0*bbs)))
                / sqrt(1 - rho^2) ) .*
           (-sxyes0) ) );
  den1 = cdfnc(- sxyes1 * bbs) - ( cdfn(- dxyes1 * bbd) -
       cdfbvn( (- dxyes1 * bbd), (- sxyes1 * bbs), rho) );
  hbbd = sumc( (-1/den1) .*
           cdfnc( ( -sxyes1*bbs - (rho * (-dxyes1*bbd)))
                / sqrt(1 - rho^2) ) .*
           pdfn( -dxyes1*bbd ) .* (-dxyes1) );
  hbbs = sumc( (1/den1) .*
           ( - pdfn(- sxyes1*bbs) + pdfn( -sxyes1*bbs ) .*
            cdfn( ( -dxyes1*bbd - (rho * (-sxyes1*bbs)))
                / sqrt(1 - rho^2) ) ) .* (-sxyes1) );
  g_bbs = bbs;
  g_bbd = bbd;
  g_rho = rho;
  bvr0 = gradp(&bvn0,g_rho);
  bvr1 = gradp(&bvn1,g_rho);
  grho = sumc( (1/den0) .* (- bvr0) );
  hrho = sumc( (1/den1) .* bvr1 );

  k = ((fb1~fb2~fb3) + gbbs'+ hbbs')
    ~ (gbbd + hbbd)' ~ (grho + hrho);

  retp(-k);
```

```
endp;


proc bvn0(r);
    local r0;
    r0 = r;
    retp( cdfbvn( (- dxyes0 * g_bbd),
                  (- sxyes0 * g_bbs), r0) );
endp;

proc bvn1(r);
    local r1;
    r1 = r;
    retp( cdfbvn( (- dxyes1 * g_bbd),
                  (- sxyes1 * g_bbs), r1) );
endp;

iter = iter + 1;
endo;

system;
```

BIBLIOGRAPHY

BIBLIOGRAPHY

Bloom, D. E., and M. R. Killingsworth (1985), "Correcting for Truncation Bias Caused by A Latent Truncation Variable," Journal of Econometrics, 27(1): 131-135.

Bockstael, N. E., Strand, I. E., McConnell, K. E., and F. Arsanjani (1990), "Sample Selection Bias in the Estimation of Recreation Demand Function: An Application to Sportfishing," Land Economics, 66(1): 40-49.

Borjas, G. J. (1987), "Self-Selection and the Earning of Immigrants," American Economic Review, 77(4): 531-553.

Bowker, J. M., and J. R. Stoll (1988), "Use of Dichotomous Choice Nonmarket Methods to Value the Whooping Crane Resource," American Journal of Agricultural Economics, 70(2): 372-381.

Brown, T. L., Dawson, C. P., Hustin, D. L., and D.J. Decker (1981), "Comments on the Importance of Late Respondent and Nonrespondent Data from Mail Surveys," Journal of Leisure Research, 13(1): 76-79.

Brown, T. L., Decker, D. J., and N. A. Connelly (1989), "Response to Mail Surveys on Resource-based Recreation Topics: A Behavioral Model and an Empirical Analysis," Leisure Sciences, 11(?): 99-110

Cameron, T. A. (1988), "A New Paradigm for Valuing Non-market Goods Using Referendum Data: Maximum Likelihood Estimation by Censored Logistic Regression," Journal of Environmental Economics and Management, 15(3): 355-379.

Cameron, T. A., and M. D. James (1987), "Efficient Estimation Methods For "Closed-ended" Contingent Valuation Surveys," The Review of Economics and Statistics, 69(2): 269-276.

Dhrymes, P. J. (1970), Econometrics, New York, NY: Harper & Row.

Edwards, S. F., and G. D. Anderson (1987), "Overlooked Biases in Contingent Valuation Surveys: Some Considerations," Land Economics, 63(2): 168-178.

Fuller, W. A. (1987), Measurement Error Models, New York, NY: John Wiley & Sons.

Goldberger, A. S. (1981), "Linear Regression After Selection," Journal of Econometrics, 15(3): 357-366.

Goldberger, A. S. (1991), A Course in Econometrics, Cambridge, MA: Harvard University Press.

Goyder, C. J. (1982), "Further Evidence on Factors Affecting Response Rates to Mailed Questionnaires," American Sociological Review, 47(4): 550-553.

Green, E. K. (1991), "Reluctant Respondents: Differences Between Early, Late, and Nonresponders to a Mail Survey," Journal of Experimental Education, 59(3): 268-276.

Green, E. K., and R. F. Kvidahl (1989), "Personalization and Offers of Results: Effects on Response Rates," Journal of Experimental Education, 57(3): 263-270.

Green, E. K., and S. F. Stager (1986), "The Effects of Personalization, Sex, Locale, and Level Taught on Educators' Responses to a Mail Survey," Journal of Experimental Education, 54(4): 203-206.

Greene, W. H. (1981), "Sample Selection Bias as a Specification Error: Comment," Econometrica, 49(3): 795-798.

Greene, W. H. (1990), Econometric Analysis, New York, NY: Macmillan Publishing Company.

Hauseman, J. A., and D. A. Wise (1977), "Social Experimentation, Truncated Distributions, and Efficient Estimation," Econometrica, 45(4): 919-938.

Hauseman, J. A., and D. A. Wise (1981), "Stratification on Endogenous Variables and Estimation: The Gary Income Maintenance Experiment," in Manski, C.F., and D. McFadden (eds.), Structural Analysis of Discrete Data: With Econometric Applications, 51-111, Cambridge, MA: MIT Press.

Heckman, J. J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," Annals of Economic and Social Measurement, 5(4): 475-492.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," Econometrica, 47(1): 153-161.

Heckman, J. J., and G. L. Sedlacek (1985), "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market," Journal of Political Economy, 93(6): 1077-1125.

Heckman, J. J., and G. L. Sedlacek (1990), "Self-Selection and the Distribution of Hourly Wages," Journal of Labor Economics, 8(1): s329-s363.

Hoehn, J. P. and J. B. Loomis (1993), "Substitution Effects in the Valuation of Multiple Environmental Programs," Journal of Environmental Economics and Management, 25(1): 56-75.

Kanuk, L., and C. Berenson (1975), "Mail Surveys and Response Rate: A Literature Review," Journal of Marketing Research, 12(4): 440-453.

Lee, L. F. (1984), "Tests for Bivariate Normal Distribution in Econometric Models with Selectivity," Econometrica, 52(4): 843-863.

Little, R. J. A. (1985), "A Note About Models for Selectivity Bias," Econometrica, 53(6): 1469-1474.

Little, R. J. A., and D. B. Rubin (1987), Statistical Analysis with Missing Data, New York, NY: John Wiley & Sons.

Loomis, J. B. (1987), "Expanding Contingent Value Sample Estimates to Aggregate Benefit Estimates: Current Practices and Proposed Solutions," Land Economics, 63(4): 396-402.

Maddala, G. S. (1983), Limited-dependent and Qualitative Variables in Econometrics, New York, NY: Cambridge University Press.

McConnell, K. E. (1990), "Models for Referendum Data: The Structure of Discrete Choice Models for Contingent Valuation," Journal of Environmental Economics and Management, 18(1): 19-34.

Mitchell, R. C., and R. T. Carson (1989), Using Surveys to Value Public Goods: The Contingent Valuation Method, Washington, D.C.: Resources for the Future.

Muthén, B., and K. G. Jöreskog (1983), "Selectivity Problems in Quasi-experimental Study." Evaluation Review, 7(2): 139-174.

Nelson, F. D. (1977), "Censored Regression Models with Unobserved Stochastic Censoring Thresholds," Journal of Econometrics, 6(3): 309-327.

Olsen, R. J. (1980), "A Least Squares Correction For Selectivity Bias," Econometrica, 48(7): 1815-1820.

Ong, P. M., Holt, S., Skumatz, L. A., and R. S. Barnes (1988), "Nonresponse in Residential Energy Surveys: Systematic Patterns and Implications for End-Use Models," Energy Journal, 9(2): 137-151.

Pudney, S. (1989), Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes, Cambridge, MA:Basil Blackwell.

Randall, A. (1987), Resource Economics, 2nd Ed., New York, NY: John Wiley & Sons.

Rubin, D. B. (1987), Multiple Imputation for Nonresponse in Surveys, New York, NY: John Wiley & Sons.

Shaw, D. (1988), "On-site Samples' Regression: Problems of Non-negative Integers, Truncation, and Endogenous Stratification," Journal of Econometrics, 37(2): 211-223.

Smith, V. K. (1988), "Selection and Recreation Demand," American Journal of Agricultural Economics, 70(1): 29-36.

van Ravenswaay, E. O. and J. P. Hoehn (1991a), "Contingent Valuation and Food Safety: The Case of Pesticide Residues in Food," Staff Paper No. 91-13, Department of Agricultural Economics, Michigan State University.

van Ravenswaay, E. O. and J. P. Hoehn (1991b), "Consumer Willingness to Pay for Reducing Pesticide Residues in Food: Results of a Nationwide Survey," Staff Paper No. 91-18, Department of Agricultural Economics, Michigan State University.

Walsh, R. G., Loomis, J. B., and R. A. Gillman (1984), "Valuing Options, Existence, and Bequest Demands for Wilderness," Land Economics, 60(1): 14-29.

Whitehead, J. C. (1991), "Environmental Interest Group Behavior and Self-selection Bias in Contingent Valuation Mail Surveys," Growth and Change, 22(1): 10-21.

Willis, R. J., and S. Rosen (1979), "Education and Self-Selection," Journal of Political Economy, 87(5): s2-s35.

Yatchew, A., and Z. Griliches (1985), "Specification Error in Probit Models," The Review of Economics and Statistics, 67(1): 134-139.