



133
245
THS



3 1293 01701 4741

This is to certify that the

dissertation entitled

Essays on Sample Selection, Self Selection
and Model Selection

presented by

Ai-Chi Hsu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Offey M. Walbridge
Major professor

Date June 24, 1998

PLACE IN RETURN BOX to remove this checkout from your record.
 TO AVOID FINES return on or before date due.
 MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
JUN 18 2001	FEB 08 2002	
APR 19 2001		
070701		
NOV 24 2003		

**ESSAYS ON SAMPLE SELECTION, SELF SELECTION
AND MODEL SELECTION**

By

Ai-Chi Hsu

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Economics

1998

ABSTRACT

ESSAYS ON SAMPLE SELECTION, SELF SELECTION AND MODEL SELECTION

By

Ai-Chi Hsu

This dissertation analyzes several different issues concerning selection corrections with a censored selection variable, including sample selection bias corrections, self-selection bias corrections, and some model selection issues. Several methods are developed to correct sample and self-selection biases. Model selection for hurdle-type models is also discussed.

Chapter 1 develops a sample selection bias correction procedure when the endogenous explanatory variable appearing in the structural equation is also the variable determining selection. Assuming that the selection variable follows a standard Tobit, a simple two-step estimator is consistent and asymptotic normal. The method is applied to a popular data set from the labor economics literature. Some comparisons also are made for different estimation approaches.

Chapter 2 develops a method to simultaneously deal with self-selection and sample selection problems with a Tobit selection equation and a roughly continuous endogenous explanatory variable, as in Garen (1984). This method is applied to two different data sets from the labor economics literature.

Chapter 3 studies model selection for hurdle models that are often used as extensions of the Tobit model. Two competing non-nested models are considered. One is Cragg's (1971) truncated normal model that assumes the distribution of the dependent variable

follows a normal distribution and truncated at zero. The other one is log-normal model that assumes the distribution of the dependent variable follows a log-normal distribution. To select one model rather than the other, Vuong's (1989) approach is applied to see which model fits data better. The simulation results show that Vuong's test has reasonable power for choosing the correct model.

Dedicated to My Parents and Brother

ACKNOWLEDGEMENTS

First and foremost, I would like to say ‘thank you!’ to my dissertation advisor Professor Jeffrey Wooldridge. Professor Wooldridge shared his professional knowledge and valuable time with me. This dissertation would never have been possible without his help and support. I also thank other committee members, Professor Robert Lalonde and Professor Stephen Woodbury for their helpful comments on my thesis, especially chapter 2.

There are some friends and fellow graduate students to whom I also want to say thanks. Thanks to Hui-Wen Shih, Cheng-Ping Cheng, Te-Fen Lo, I-Jung Tsai, and the other graduate students I worked with.

I am grateful to my parents and my brother for their support throughout my graduate study. Thanks to them again.

TABLE OF CONTENTS

LIST OF TABLES.....	vi
INTRODUCTION	1
 CHAPTER 1	
SELECTION CORRECTION WITH ENDOGENOUS EXPLANATORY VARIABLES AND A TOBIT SELECTION EQUATION	5
I. Introduction	5
II. The Basic Model.....	10
III. Empirical Application.....	17
III.1 Estimating the Structural Wage Equation.....	17
III.2 Estimating the Labor Supply Equation.....	22
IV. An Extension of the Model.....	23
V. Models with Additional Endogenous Explanatory Variables.....	24
VI. Conclusion.....	26
 CHAPTER 2	
SELECTION CORRECTIONS WHEN BOTH SELF-SELECTION BIASES AND SAMPLE SELECTION BIASES ARE PRESENT.....	33
I. Introduction.....	33
II. A Random Coefficient Model.....	36
III. Garen's Model with Sample Selection.....	39
IV. Empirical Examples.....	42
V. The Consistency of Procedure 2.2.....	49
VI. Conclusion.....	52
 CHAPTER 3	
MODEL SELECTION TESTS FOR TWO-PART MODELS.....	63
I. Introduction.....	63
II. Basic Framework.....	68
III. Empirical Result.....	73
(i) The Simulation Results.....	74
(ii) The real Data Sets Results.....	77
IV. Conclusion.....	78

CHAPTER 4
CONCLUSION..... 82
LIST OF REFERENCES..... 85

LIST OF TABLES

Chapter 1

SELECTION CORRECTION WITH ENDOGENOUS EXPLANATORY VARIABLES AND A TOBIT SELECTION EQUATION

Table 1.1: The Descriptive Statistic of NLS Data Set.....	28
Table 1.2: The Estimation of Different Specifications.....	29
Table 1.3: Two step and Moffitt's MLE estimation.....	30
Table 1.4: Tobit Estimation for the Labor Supply Equation.....	31
Table 1.5: Two Step Estimation with assumption relaxed.....	32

Chapter 2

SELECTION CORRECTIONS WHEN BOTH SELF-SELECTION BIASES AND SAMPLE SELECTION BIASES ARE PRESENT

Table 2.1: The Descriptive Statistics for PSID Data Set.....	56
Table 2.2: The Estimation of Different Specifications (PSID Data Set).....	57
Table 2.3: Joint tests for self-selection and Sample selection biases (PSID data set).....	58
Table 2.4: Descriptive Statistics for CPS Data Set.....	59
Table 2.5: The Estimation of Different Specifications (CPS Data Set).....	60
Table 2.6: Joint tests for self-selection and Sample selection biases (CPS data set).....	61
Table 2.7: The Estimation of Different Specifications (PSID Data Set without Parents' Education as Instruments).....	62

Chapter 3

ALTERNATIVE MODELS SELECTION FOR TOBIT SPECIFICATION

Table 3.1: Combinations of the Simulation.....	79
Table 3.2: The Rejection Rate (True Model: Log Normal Model).....	80
Table 3.3: The Rejection Rate (True Model: Truncate Normal Model).....	81

INTRODUCTION

This dissertation analyzes several different topics for selection corrections with a censored selection variable, including sample selection bias corrections, self-selection bias corrections, and some model selection issues.

It is quite normal in econometrics to assume that a random sample is available from the underlying population of interest. However, this is not always the case. Sometimes due to the way economic data are collected and economic behavior by the individuals being sampled, a nonrandom sample can be generated. By “sample selection” I mean cases where certain variables cannot be observed for a subset of the population. Thus, sample selection concerns data availability.

Self-selection refers to any situation when one or more explanatory variables correlated with unobservable factors affecting the outcome equation. An example is education in a wage equation, where education is correlated with unobserved ability that affects wage too.

The distinction between self-selection and sample selection is not always sharp. For example, people can self-select into the workforce, which leads to a data observability problem (we do not observe the wage offer for those out of the workforce). We will treat this as a sample selection problem. The labels we give to these econometric problems are not crucial, but it is useful to have a way of distinguishing endogeneity of explanatory variables from missing data problems.

In this thesis I propose methods to correct for possible sample selection biases. I study the case when the variable determining selection follows a reduced form Tobit

model. Usually, the sample selection issue is appeared in the context of binary selection information: either we observe the data or do not. But in some cases more information is available. A leading example is estimating a wage offer equation for working age adults. Rather than work or not, the data for working hours is also available. It is possible to improve the estimation using this extra information.

As we know, Heckman's (1976) method for correcting potential sample selection biases is widely used in empirical studies. However, the method developed in this thesis has advantages over Heckman's method: the identification is easier, handling endogenous explanatory variables is easier.

In chapter 1 I develop a sample selection bias correction procedure when the endogenous variable, especially the Tobit selection variable, is an explanatory variable in the structural equation. The procedure can be used in a variety of studies. For example, it can be used to estimate a wage offer equation for a certain group of people (such as age over 25) which we use as an application in chapter 1. The working hour is the selection variable in this case. The working hour also appears as an explanatory variable in the structural equation. There is self-selection component to this example: people self-select into employment, so whether we observe wage offer or not depends on individual's decision. As we said, whether we call example like this sample selection or self-selection is not a very important issue. However, a typical self-selection issue in economics has some features that distinguish it from the sample unobservable problem.

Chapter 2 continues the reasoning in chapter 1 and develops a procedure that can handle both self-selection and sample selection problems at the same time to get a consistent estimation for the parameters in a structural equation. The problem of self-

selection often occurs when evaluating various programs such as job training, welfare, or the returns to the education using nonexperimental data. Generally, the issue is that individuals decide whether to participate in a program or to receive some treatment. If the participation decision is related to factors that affect the outcome variable, ignoring the determinates of this decision leads to biased estimates of program impacts. We consider the case that allows the endogenous explanatory variable interacted with an unobservable. The reasoning of the procedure is similar as Garen's (1984) procedure, which he used to estimate the return to education when education and unobserved ability interact. However, we extend Garen's work by considering the sample selection bias.

As we know, hurdle models are more flexible alternatives to Tobit. Chapter 3 studies model selection for hurdle or two-tier model. Two competing non-nested models are considered. One is Cragg's truncated normal model, the other is a two-part log-normal model.

For choosing between nonnested models, two very different approaches have been proposed. The earliest is based on the work of Cox(1961, 1962) who derived specification tests that use information about a specific alternative and test whether the null can predict the performance of the alternative. This approach assumes that one of the competing models is the true model and test the hypothesis based on this assumption.

The other approach is what Vuong(1989) terms a "model selection approach." Vuong bases tests of nonnested alternatives on an estimate of the Kullback-Leibler (1951) Information Criterion (KLIC), which measures the distance between two distributions relative to the true distribution. Compared with Cox's approach, Vuong's approach does not assume that one of the models is correct under H_0 . The null hypothesis is that the

models fit the data equally well, and the alternative is that one model fits better. If a model is correctly specified then, asymptotically, it produces the best fit.

I apply Vuong's (1989) approach to see which hurdle model fits data better. A small simulation study is applied to generated data sets for both the truncated normal distribution and the log-normal distribution to see if Vuong's procedure has power for picking the true model. I also apply the test to three labor data sets.

Chapter 4 contains concluding comments and directions for future research.

Chapter 1

SELECTION CORRECTION WITH ENDOGENOUS EXPLANATORY VARIABLES AND A TOBIT SELECTION EQUATION

I. Introduction

The problem of nonrandom sampling occurs frequently in econometrics. Sample selection problems can arise when a random sample is not available from the underlying population of interest. Due to the way economic data are collected, sometimes we can obtain only a nonrandom sample from the population. If the sample does not represent the population of interest, sample selection biases may arise. As an example, suppose we are interested in estimating a wage offer equation for women over age 30. By definition, this equation is supposed to represent all women over age 30, whether or not a woman is actually working at the time of the survey. Since we can only observe the wage offer for working women, we actually select our sample on this basis. The sample selection problem arises because data on a key variable, wage, is available only for a clearly defined subset of the population. This is an example of *incidental truncation* (of the wage offer, in this example), because wage is missing due to the outcome of labor force participation.

One can easily get confused about the distinction between sample selection and self-selection. Consider the above example without the incidental truncation problem. The population is working women, so we do not have a sample selection problem. However, education level might be correlated with unobserved characteristics, such as “ability,” that also have a direct effect on the wage. In other words, people ‘self-select’ their education

levels. This makes self-selection problem and sample selection problem different in the nature. In general, the terms “sample selection” and “self-selection” are used to make a clear distinguish for the nature of the selection problems. Sample selection refers to the data observation problem when there are selected samples; self-selection refers to the endogeneity of the explanatory variables when the explanatory is considered to be correlated with the error term in the structural equation. There is no strict separation between these problems. Self-selection could be the cause of the sample selection problem. As in our wage offer example, it is true we do not observe the wage offer for people who do not work, so it is a sample selection problem. However, one can also say that whether people work or not comes from their “self selection.”

It is fairly well known that in a linear model with endogenous explanatory variables, if selection rule is based on some exogenous variables, estimation of the population model by two stage least square (2SLS hereafter) using the selected sample is consistent and asymptotically normal (see, for example, Wooldridge (1996)). If the selection rule is based on endogenous variables, applying 2SLS to estimate the population model using the selected sample will not generally be consistent. Therefore, it is important to have methods for testing and correcting for sample selection with endogenous explanatory variables.

The purpose of this chapter is to derive and apply an alternative method to test and correct sample selection bias while estimating essentially the Type III Tobit model¹ with endogenous explanatory variables, concentrating on the core where the selection variable appear in the structural equation. This makes our model essentially similar to Nelson and

Olson (1978) and becomes a special case of the Type IV Tobit model, which is an extension of Type III Tobit. However, we make fewer assumptions and obtain simple two-step estimators.

A standard Type III Tobit model consists two equations. One is the structural equation, whose coefficients we want to estimate the coefficients consistently. The other is the selection equation, which takes a Tobit equation form. Unlike the Type II Tobit model, the Type III Tobit model's selection equation includes not just a binary variable (probit equation). It includes more information in the selection equation. Heckman (1976) is the pioneer of Type II Tobit model. Heckman's two step method for correcting sample selection bias for his labor-supply model by including inverse Mill's ratio into structural equation has been widely used. The procedure was further extended to a wide class of models by Lee (1978). The drawback to Heckman's procedure is that, if there is not a good exclusion restriction in the structural model, the estimators can be very imprecise in finite sample. Full maximum likelihood methods or semi-parametric method could be applied to Type II Tobit model estimation. However, as discussed by Wooldridge (1996), none of them are satisfying. If there is more information available in the selection equation, such as Type III Tobit form, the additional information may be used to get more precise estimators.

The traditional method to estimate Type III and IV Tobit model is maximum likelihood, which makes full parametric assumptions; see Amemiya (1985 section 10.8). To detect if there exists sample selection bias, Vella (1992) extended the testing procedures proposed by Heckman (1979) and Vella (1993) to construct a t-test on a

¹ See Amemiya (1985, section 10.8)

constructed variable in an auxiliary equation. However, Vella did not prove the consistency of the parameters in the structural equation using his method. In the case of a Tobit selection equation, Wooldridge (1996) showed that adding the Tobit residuals can produce consistent estimators in the structural equation. Wooldridge relaxed the basic Tobit III model assumption, which needed by MLE, that the distribution of the error terms in both the structural equation and the selection equation are bivariate normal. Instead, he proposed a weaker assumption that only the error term in the selection equation is normal distributed and the two error terms have a linear relationship. Under these modifications of the Type III Tobit model, he proposed a multi-step procedure to estimate the model. At the same time, he also proved that the testing procedure proposed by Vella can not only be a test but a correction for sample selection bias in cross section linear model case. In other word, he proved the consistency of the parameters in the structural equation after the constructed variable is added as a regressor.

Unlike MLE's computational burden, the multiple-step procedure is much easier to implement, even for models with endogenous explanatory variables (the case we are going to discuss in this paper). It also requires weaker assumptions than MLE. In other word, it is more robust. Furthermore, it can provide a simple test for the sample selection bias and correct the problem at the same time if biases do exist.

In this chapter, the basic approach used is a simpler two-step method that extends a method proposed by Wooldridge (1996). Wooldridge covers the case of Type III Tobit model. I focus on the case that the Type III Tobit model with endogenous explanatory variables, especially the case where the dependent variable of the selection equation itself turns out to be an explanatory variable in the structural equation (a special case of Type

IV Tobit). To consistently estimate our model, we make some adjustment of the Type III Tobit model by putting the dependent variable of the selection equation into the right hand side of structural equation as an explanatory variable (a special case of Type IV Tobit) and correcting the sample selection bias and endogeneity of the selection variable at the same time. We even extend the two-step procedure to three-step procedure in order to estimate the selection equation consistently. We use the same data set of Moffitt (1984), which contains the information on married women in 1972 wave of the National Longitudinal Survey of Older Women. Moffitt used a maximum likelihood method to estimate the labor supply function in his paper. Comparing the method he used, the method we use here is easier to compute and requires fewer prior assumptions, which means the estimation procedure we use is more robust.

The results of the estimation show that 1): there does exist sample selection bias that cannot be ignored. 2): the effect of hours on $\log(\text{wage})$ appears to be linear. 3): the returns of education appears to be overestimated when the sample selection problem is not accounted for. 4): the wage effect is positive and significant in the labor supply equation. 5): the wage elasticity of the labor supply is less when it is conditional on working hours are positive than without conditions.

The basic econometric model is in section II. The empirical results and the comparisons are proposed in section III. We relax the assumption of a linear relationship between the error terms in structure equation and selection equation assumed in section IV. In section V we extend the basic model to allow additional endogenous explanatory variables in the structural equation. Section VI is the conclusion.

II. The Basic Model

Consider the following structural model:

$$y_1 = \beta y_2^* + z_1 \gamma + u_1 \quad (1.1)$$

$$y_2^* = \alpha_0 y_1 + z_2 \delta + u_2 \quad (1.2)$$

$$y_2 = \max(0, y_2^*) \quad (1.3)$$

where y_1, y_2 are endogenous variables and y_1 is observed only when $y_2 > 0$; z_1, z_2 are exogenous variables and are always observed, z_1 is a $1 \times K_1$ vector with $z_{11} \equiv 1$ and z_2 is a $1 \times K_2$ vector with $z_{21} \equiv 1$; u_1, u_2 are structural errors; y_2^* is the part of y_2 where $y_2 > 0$.

This model basically is in the category of the Tobit selection model with endogenous explanatory variables. The basic framework is similar to Nelson and Olson (1978). However, Nelson-Olson assumes y_1 is always observed so there is no sample selection problem, which is different from our specification. We use y_2^* instead of y_2 as explanatory variable in (1.1) so that the reduced form of equation (1.2) can be derived. The reduced form of equation (1.2) can be written as (1.3) so the general model becomes (1.1), (1.3) and

$$y_2^* = z\pi_2 + v_2 \quad (1.4)$$

We can combine (1.3) and (1.4) to obtain a reduced form selection equation:

$$y_2 = \max(0, z\pi_2 + v_2) \quad (1.5)$$

The $1 \times K$ vector z contains nonredundant elements z_1 and z_2 . This model is similar as Type III Tobit model except we include y_2^* in (1.1) as explanatory variable.

We consider a general linear model where selection is determined by the instrumental variables. The population model of interest is

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k + u = x\beta + u,$$

where $x_1 \equiv 1$. One or more of the elements of x can be correlated with u . We also assume the availability of a $1 \times L$ vector z , $L \geq k$, such that

$$E(u | z) = 0.$$

Naturally, any exogenous elements in x are included in z . Under the rank condition for identification, $\text{rank } E(z'x) = k$, β could be consistently estimated by two stage least squares if we have a random sample from the population. However, if we can observe data on (x, y, z) for only a subset of the population, consistency depends on the nature of the selection rule.

Let s be a binary selection indicator such that $s = 1$ if (x, y, z) is observed and $s = 0$ otherwise. Assume that $s = h(z)$ for some known function $h(\cdot)$, so that selection is based entirely on the exogenous instrument variables. Let (x_i, y_i, z_i) , $i=1, 2, \dots, N$, be a random sample from the population, and let $s_i = h(z_i)$. Then the 2SLS estimator using the selected sample can be written as

$$\hat{\beta} = \left[\left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right)' \left(N^{-1} \sum_{i=1}^N s_i z_i' z_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right) \right]^{-1} \cdot \left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right)' \left(N^{-1} \sum_{i=1}^N s_i z_i' z_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N s_i z_i' y_i \right).$$

Substituting $y_i = x_i\beta + u_i$ gives

$$\hat{\beta} = \beta + \left[\left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right)' \left(N^{-1} \sum_{i=1}^N s_i z_i' z_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right) \right]^{-1} \cdot \left[\left(N^{-1} \sum_{i=1}^N s_i z_i' x_i \right)' \left(N^{-1} \sum_{i=1}^N s_i z_i' z_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N s_i z_i' u_i \right) \right].$$

Since we assume $E(u | z) = 0$ and $s_i z_i$ is some function of z_i , $E(s_i z_i' u_i) = 0$ by iterated expectations. According to the law of large numbers, $\text{plim } \hat{\beta} = \beta$. More details are given in Wooldridge (1996).

In our setting, selection is determined by a censored selection variable. Let v be a zero mean random variable such that (z, v) determines the sample selection: $s = h(z, v)$ and (x, y, z, v) is observed when $s = 1$. Assuming that (u, v) is independent of z , and that $E(u | v) = \xi v$, we can write

$$y = x\beta + \xi v + e, \tag{1.6}$$

where $E(e | z, v) = 0$. If we use 2SLS on the selected sample to estimate (1.1), we will not get a consistent estimator of β . But we can estimate (1.6) by 2SLS using instruments (z, v) .

We can apply this result to model (1.1) and (1.5) to estimate β and γ . To test and correct the possible sample selection bias when estimating in (1.1), we need a few assumptions:

Assumptions 1.1

- (i) (z, y_2) is always observed but y_1 is observed only when $y_2 > 0$.

(ii) (u_1, v_2) is zero mean and independent of z .

(iii) $v_2 \sim \text{Normal}(0, \tau_2^2)$

(iv) $E(u_1 | v_2) = \xi v_2$

An important assumption in the Type III Tobit model is the bivariate normality assumption so that the model can be estimated by MLE. However, here we only need Assumptions 1.1 (iv) instead of the bivariate normality of (u_1, v_2) . This allows a fairly broad range of distributions for u_1 . Under Assumptions 1.1 (ii) and (iv) we can write

$$u_1 = \xi v_2 + e \quad (1.7)$$

where e is a disturbance such that $E(e | z, v_2) = 0$. Plugging (1.7) into (1.1) gives

$$y_1 = \beta y_2^* + z_1 \gamma + \xi v_2 + e. \quad (1.8)$$

Here the test of $H_0 : \xi = 0$ is a valid test for sample selection problem, as it tests whether u_1 and v_2 are uncorrected. If H_0 is not rejected, y_2^* is exogenous in the structural equation (1.1) and selection is based on exogenous variables (since z is exogenous). Thus, OLS on the selected sample works in that case. On the other hand, y_2^* is endogenous in the structural equation (1.1) if $\xi \neq 0$. However, since y_2^* depends on (z, v_2) , it is exogenous in equation (1.8). In fact,

$$E(y_1 | z, v_2) = \beta y_2^* + z_1 \gamma + \xi v_2 \quad (1.9)$$

Further, the expectation is the same if we condition also on $v_2 > -z\pi_2$. From Wooldridge (1996) Theorem 2.1, it follows that OLS estimation of (1.9) using the selected sample will be consistent (take $x = (y_2^*, z_1, v_2)$). Because v_2 is not observed, it must be estimated.

This leads to the following procedure

PROCEDURE 1.1:

(i) Estimate

$$y_{i2} = \max(0, z_i \pi_2 + v_{i2}), \quad v_{i2} | z_i \sim \text{Normal}(0, \tau_2^2)$$

by standard Tobit MLE using all N observations.

(ii) Obtain the residuals from step 1 as \hat{v}_{i2} for $y_{i2} > 0$ ($i = 1, 2, \dots, N_1$) by defining

$$\hat{v}_{i2} = y_{i2} - z_i \hat{\pi}_2, \quad i = 1, 2, \dots, N_1.$$

(iii) Regress y_{i1} on y_{i2} , z_{i1} , and \hat{v}_{i2} , $i = 1, 2, \dots, N_1$

Since we have consistent first-stage estimates, replacing v_2 with estimates does not affect consistency in the second stage. The correction of standard errors due to the generated regressors problem is discussed in Wooldridge (1996). Wooldridge's formula can be applied here because \hat{v}_2 is the only generated regressor.

It should be noticed that there are two separate features of our model. First, there exists an endogeneity problem for y_2^* as an explanatory variable in the structural equation. Second, there exists sample selection problem because we can only observe y_1 when $y_2 > 0$. By including v_2 as an additional regressor in the structural equation, the two problems are solved at the same time. The endogeneity of y_2 actually is the source of the sample selection problem in our case. If y_2 is exogenous, then u_1 and v_2 are not correlated with each other. The sample selection problem does not exist.

In addition to β and γ , we may want to estimate α_0 and δ . Let

$$\hat{y}_{i1} \equiv \hat{\beta} \hat{y}_{i2}^* + z_{i1} \hat{\gamma} \quad \text{for all observations } N, \text{ where } \hat{y}_{i2}^* = z_i \hat{\pi}_2 \text{ is the predicted value of the}$$

Tobit model in the Procedure 1.1 step (i). Estimate the model

$$y_{i2} = \max (0, \alpha_0 \hat{y}_{i1} + z_{i2} \delta + error_2) \quad (1.10)$$

by standard Tobit. Then we can get consistent estimators for α_0 and δ also. Since the relationship between explanatory variables and dependent variable is nonlinear, interest lies in $E(y_{i2}|y_{i1}, z_{i2})$ and $E(y_{i2}|y_{i1}, z_{i2}; y_{i2} > 0)$. If we let $x\theta \equiv \alpha_0 \hat{y}_{i1} + z_{i2} \delta$, we can apply the formulas provided by Wooldridge (1998) directly. We briefly outline the reasoning of the formulas in the following,

To obtain $E(y_{i2}|x)$, first use the law of iterated expectations:

$$\begin{aligned} E(y_{i2}|x) &= P(y_{i2} = 0|x) \cdot 0 + P(y_{i2} > 0|x) \cdot E(y_{i2}|x, y_{i2} > 0) \\ &= P(y_{i2} > 0|x) \cdot E(y_{i2}|x, y_{i2} > 0) \end{aligned} \quad (1.11)$$

the first term $P(y_{i2} > 0|x)$ is easy to be derived,

$$\begin{aligned} P(y_{i2} > 0|x) &= P(u_2 > -x\theta|x) \\ &= P\left(\frac{u_2}{\sigma} > -x\theta/\sigma\right) = \Phi(x\theta/\sigma) \end{aligned} \quad (1.12)$$

We can also derive that

$$E(y_{i2} | x, y_{i2} > 0) = x\theta + E(u_2 > -x\theta) = x\theta + \sigma \frac{\phi(x\theta/\sigma)}{\Phi(x\theta/\sigma)}$$

So we can conclude that

$$E(y_{i2}|x, y_{i2} > 0) = x\theta + \sigma \frac{\phi(x\theta/\sigma)}{\Phi(x\theta/\sigma)} \quad (1.13)$$

$$E(y_{i2}|x) = \Phi(x\theta/\sigma)x\theta + \sigma\phi(x\theta/\sigma) \quad (1.14)$$

where σ is standard error, $\phi(\cdot)$ is the standard normal pdf, and $\Phi(\cdot)$ is the standard normal cdf. The partial effect of variable x_j with respect to (1.13) and (1.14) are:

$$\begin{aligned}\frac{\partial E(y_{i2}|x; y_{i2} > 0)}{\partial x_j} &= \theta_j + \theta_j \cdot \frac{d\lambda}{d\sigma}(x\theta/\sigma) \\ &= \theta_j [1 - \lambda(x\theta/\sigma) \{x\theta/\sigma + \lambda(x\theta/\sigma)\}] \end{aligned} \quad (1.15)$$

where $\lambda = \frac{\phi(\cdot)}{\Phi(\cdot)}$, and

$$\frac{\partial E(y_{i2}|x)}{\partial x_j} = \frac{\partial P(y_{i2} > 0 | x)}{\partial x_j} \cdot E(y_{i2} | x, y_{i2} > 0) + P(y_{i2} > 0 | x) \cdot \frac{\partial E(y_{i2} | x, y_{i2} > 0)}{\partial x_j}$$

since we already know that $P(y_{i2} > 0 | x) = \Phi(x\theta/\sigma)$ and $\frac{\partial P(y_{i2} > 0 | x)}{\partial x_j} = \frac{\theta_j \phi(\cdot)}{\sigma}$, we

can derive that

$$\begin{aligned}\frac{\partial E(y_{i2}|x)}{\partial x_j} &= \frac{\theta_j}{\sigma} \phi(x\theta/\sigma) \cdot E(y_{i2} | x, y_{i2} > 0) + \Phi(x\theta/\sigma) \cdot \frac{\partial E(y_{i2} | x, y_{i2} > 0)}{\partial x_j} \\ &\dots\dots\dots \end{aligned} \quad (1.16)$$

The elasticity of y_{i2} with respect to x_j , conditional on $y_{i2} > 0$, is

$$\frac{\partial E(y_{i2} | x, y_{i2} > 0)}{\partial x_j} \cdot \frac{x_j}{E(y_{i2} | x, y_{i2} > 0)} \quad (1.17)$$

and the elasticity of y_{i2} with respect to x_j , without conditional on $y_{i2} > 0$, is

$$\frac{\partial E(y_{i2} | x)}{\partial x_j} \cdot \frac{x_j}{E(y_{i2} | x)} \quad (1.18)$$

In section III we apply the three-step method we derived so far to estimate wage equation

and the correspondent labor supply equation. We also estimate the wage elasticity of labor supply.

We can use the same argument to put y_2^{*2} in the structural equation of our model.

The sample selection correction version of the structural equation becomes:

$$y_1 = \beta y_2^* + \eta y_2^{*2} + z_1\gamma + \xi v_2 + e \quad (1.19)$$

where y_2^{*2} in this equation is an exogenous explanatory variable. However, except for $\eta = 0$, (1.19), (1.2) and (1.3) can not lead us to have a reduced form selection equation like (1.5). But we can use (1.19) to test $H_0 : \eta = 0$; that is, we can test for linearity in the relationship between y_1 and y_2^*

III. Empirical Application

III.1. Estimating the Structural Wage Equation

We now apply the previous methods to consistently estimate a wage offer equation. The structural equations are:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{hours}^* + z_1\gamma + u_1 \quad (1.20)$$

$$\text{hours} = \max(0, \alpha_0 \log(\text{wage}) + z_2\delta + u_2) \quad (1.21)$$

where hours^* are “desired” hours, observed only when $\text{hours} > 0$. z_1 and z_2 are vectors of exogenous variables as defined in section II. In the notation of section II, take

$y_1 = \log(\text{wage})$ and $y_2^* = \text{hours}^*$. We observe wage only when $\text{hours} > 0$. Also, we assume Assumptions 1.1.

We can derive the reduced form of (1.21) as

$$\text{hours} = \max(0, z\pi_2 + v_2). \quad (1.22)$$

There are two reasons we put hours^* as an explanatory variable in the wage equation instead of the traditional case that does not have it included in the wage equation. First, theoretically labor is a quasi-fixed input that will generate quasi-fixed costs to the firm, which means that a lower wage will be paid at lower hours of work than at high hours of work; see Oi (1962). There have been few theoretical or empirical studies of the effect of hours worked on wage offers. Rosen (1969) is an example of an early study. Rosen (1976), Larson (1979), and Hausman (1981) also are examples. Second, Moffitt (1984) argues that working hours is an important explanatory variable that should be included in the wage equation. We apply the current methods to Moffitt's data set to see if this is the case.

From Assumptions 1.1 in section II, we know that:

$$E(\log(\text{wage}) | z, v_2) = \beta_0 + \beta_1 \text{hours}^* + z_1\gamma + \xi v_2 \quad (1.23)$$

Adding the term ξv_2 is not only for testing and correcting sample selection bias but also for testing and correcting the endogeneity of hours^* . The consistent estimators of β_1 and γ can be obtained by the following procedure:

(a). Estimate (1.19) by standard Tobit for all observations N. Save the residual of (1.19)

for $\text{hours} > 0$ as \hat{v}_2 by defining $\hat{v}_2 = \text{hours} - z\pi_2$, for $\text{hours} > 0$

(b). Regress $\log(\text{wage})$ on 1, hours^* , z_1 , and \hat{v}_2 .

Here we assume that $\text{Var}(u_1|v_2)$ is constant so that the t statistic is valid and we can get a convenient t-test of $\hat{\xi}$ for sample selection bias/endogeneity of hours^* after the regression in step (b). If it is significant, we have also correct the problem. The White (1980) heteroskedasticity-robust t-statistic can be used if $\text{Var}(u_1|v_2)$ is not constant.

We estimate the model using the same data set used by Moffitt (1984). The data set contains a cross section of women drawn from the 1972 wave of the National Longitudinal Survey (NLS) of Older Women. It includes all women in this wave who have valid data for hours of work last week, the hourly wage rate, the flow of asset income, and related characteristics. The variables included in the z_1 vector are race, age, years of schooling, and three area variables: the size of local labor force, the employment fractions in manufacturing and in government in the census region of residence. The z_2 vector includes dummies for marital status, age, race, the number of family members, the number of children in the household who are less than 6, the number of children in the household who are greater than 6. Both z_1 and z_2 are treated as exogenous explanatory variables. See Table 1.1 for more details about the descriptive statistics. In Table 1.1 the percentage of women working is 48.69%. Eighty percent of the women are married. The average age of the sample is 44.37 and the average education is 10.32 years.

To compare the differences between ordinary least square estimation and our two step procedure, we estimate the structural equation using ordinary least square on the selected sample. This corresponds to simply dropping \hat{v}_{i2} as an explanatory variable. We

also want to see if hours^2 any importance for determining wage offer by adding hours^2 as an additional explanatory variable in wage equation. The results are shown in Table 1.2.

In Table 1.2, the t-test for \hat{v}_{i2} is significant in column (2). It indicates that there does exist sample selection bias. Comparing the coefficient of educ in columns (1) and (2), that is, with and without \hat{v}_{i2} into the structural equation, the coefficient decreased from .0777 to .0594. The return of education appears to be overestimated if the sample selection bias is not corrected. The coefficient of hours^* also changes from .0036 to .0121 before and after we add the constructed variable \hat{v}_{i2} into the structural equation. This means that if we treat hours^* as an exogenous variable, the return of working hours is apparently underestimated. In column (3) we add hours^{*2} as an additional regressor in our two step procedure to see if it is important for determining wage, as found by Moffitt. However, it is not significant in our estimation.

Since Moffitt used wage instead of $\log(\text{wage})$ as the dependent variable in the wage equation, we have to make the modification in our structural equation using wage as the dependent variable in order to make a direct comparison with Moffitt's maximum likelihood results. The results are in Table 1.3.

In Table 1.3, column (1) is our two step procedure using wage as the dependent variable. The t-statistic for \hat{v}_2 is still significant as in Table 1.2, which indicate sample selection problem. The effect of education in this case is 16.06 cents. We also find that the estimate coefficient for hours is .243 and is significant. In column (2) we add hours^{*2} as additional regressor in the two step procedure. Just as in Table 1.2, we still do not find that the coefficient of hours^{*2} is significant in our estimation. The estimated coefficient for

hours also is not significant in column (2) after we add hours^{*2} . This maybe come from the multicollinearity between hours and hours^{*2} . Since Moffitt did not accountd for the sample selection problem in his study, we put the OLS estimation of the wage equation without adding \hat{v}_2 in column (3) to compare with Moffitt's estimation in column (4). We find that the estimation results are quite similar except for the coefficients of hours and hours^{*2} . The effects of education are 20 cents in both cases compared with around 16 cents if the sample selection/endogeneity problem is taken care of.

Returning to Table 1.2 for our original setting, just like Moffitt found, we find there are significant effect of hours of work on the wage. After correcting the sample selection bias, the effect is more significant.

Moffitt (1984) found a quadratic relationship between wage and hours. As discussed earlier, we can use the same framework to test for hours^2 by adding it to the structural equation. Although we put hours^* on the right hand side of the wage equation, the coefficient of \hat{v}_{i2} which represent the estimation of the correlation of u_i and \hat{v}_{i2} is still significant. This is different from what Moffitt found. In our test, the sample selection bias does exist. Also, in Moffitt's paper, he found that hours^2 has a significant negative effect on wage. After putting hours^2 in the right hand side of the structural equation, we did not find that effect is significant. In Table 1.3, the t-statistic for hours^{*2} is -.25, which is very insignificant. Thus, while the point estimate suggests a decreasing return to hours, as Moffitt found, the effect is not significant.

III.2 Estimating the Labor Supply Equation

After we estimate the structural wage equation, we can also estimate the labor supply equation too. Since we did not find the significance of hours^{*2} , we only include hours^* in the structure equation as an explanatory variable instead of hours^* and hours^{*2} . The procedure just like we derived in section II. We post the result of Tobit estimation for the labor supply equation in Table 1.4.

To derive the wage elasticity of labor supply, let $\alpha_0 \log(\text{wage}) + z_2 \delta = x\theta$. According to equation (1.13) and (1.14), the partial effects of wage with respect to labor supply with and without the condition $\text{hours} > 0$ are that

$$\frac{\partial E(\text{hours}|\log(\text{wage}), z_2; \text{hours} > 0)}{\partial \text{wage}} = \frac{\alpha_0}{\text{wage}} [1 - \lambda(x\theta/\sigma) \cdot ((x\theta/\sigma) + \lambda(x\theta/\sigma))] \quad (1.24)$$

$$\frac{\partial E(\text{hours}|\log(\text{wage}), z_2)}{\partial \text{wage}} = \frac{\alpha_0}{\text{wage} \cdot \sigma} \Phi(\cdot) (x\theta + \sigma \lambda(\cdot)) + \Phi(\cdot) \frac{\alpha_0}{\text{wage}} (1 - \lambda(\cdot)(x\theta/\sigma + \lambda(\cdot)))$$

..... (1.25)

The correspondent elasticities are:

1. The wage elasticity conditional on working hours > 0 (we call it elas1)

$$\text{elas1} = \frac{\partial E(\text{hours}|\log(\text{wage}), z_2; \text{hours} > 0)}{\partial \text{wage}} * \frac{\text{wage}}{x\theta + \sigma \lambda(\cdot)} \quad (1.26)$$

2. The wage elasticity without conditional on working hours > 0 (we call it elas2)

$$\text{elas2} = \frac{\partial E(\text{hours} | \log(\text{wage}), z_2)}{\partial \text{wage}} * \frac{\text{wage}}{\Phi(\cdot)x\theta + \sigma\phi(\cdot)} \quad (1.27)$$

Plugging in the mean values of the estimates, (1.24) equals to 3.71 and (1.25) equals to 5.31. We can also calculate the wage elasticity of labor supply for these two cases. In the case of $\text{hours} > 0$, the elasticity is 0.26. This is pretty close to the nonlinear constraint case in Moffitt's study, which is 0.21. If we do not condition on $\text{hours} > 0$, the elasticity is 0.686.

As Moffitt found, we can see in Table 1.4 the other coefficients in the labor supply equation are generally of the expected sign but usually of low significance. Married women work less, as do older women and those who have more children.

IV. An Extension of the Model

It is natural to extend our model to relax some assumptions and get a more robust estimation. Keeping all assumptions in Assumption 1.1 except (iv), we instead assume $E(u_1 | v_2) = \zeta_1 v_2 + \zeta_2 (v_2^2 - \tau_2^2)$. The conditional expectation of the structure equation becomes

$$E(y_1 | z, v_2) = -\zeta_2 \tau_2^2 + \beta y_2^* + z_1 \gamma + \xi_1 v_2 + \zeta_2 v_2^2 \quad (1.28)$$

Since we still assume the normality of v_2 , we can get a consistent estimator of π_2 from a Tobit equation. So \hat{v}_{i2} and \hat{v}_{i2}^2 can be used in the procedure 2-1 we described in section II. The test for sample selection bias/endogeneity is now the joint null $H_0: \zeta_1 = 0, \zeta_2 = 0$. The result shows in Table 1.5.

In Table 1.5, the estimated return to education is essentially unchanged when \hat{v}_2^2 is added to the structural equation as an explanatory variable. The change in the coefficient on hours is also very small. It changes from .0121 to .0122. This is not surprising because \hat{v}_2^2 is not very significant using the standard t-statistic. We did not find any significant change in the structural equation estimation after we allow for a quadratic in $E(u_1 | v_2)$.

V. Models with Additional Endogenous Explanatory Variables

We can add additional endogenous variables to the model in the section II. For simplicity, we consider the case of one additional endogenous explanatory variable added into the model. Consider the model:

$$y_1 = \beta_1 y_2^* + \beta_2 y_3 + z_1 \gamma + u_1 \quad (1.29)$$

$$y_2^* = z \pi_2 + v_2 \quad (1.30)$$

$$y_3 = z \pi_3 + v_3 \quad (1.31)$$

Here y_2^* is just as before and $y_2 = \max(0, z \pi_2 + v_2)$ is the selection equation. y_3 is an endogenous explanatory variable that can be a binary variable, a count variable, or contain both continuous and discrete characteristics. For simplicity, we assume y_3 is a scalar. The goal again is to estimate the structural parameters in (1.29).

Assumptions 1.2

- (i) (z, y_2) is always observed and (y_1, y_3) is observed when $y_2 > 0$.
- (ii) (u_1, v_2) is independent of z .
- (iii) $v_2 \sim \text{Normal}(0, \tau_2^2)$.

(iv) $E(u_1|v_2) = \xi v_2$.

(v) $z\pi_3 = z_1\pi_{31} + z_2\pi_{32}$, $\pi_{32} \neq 0$.

There are two cases of interest in the setup of (1.29) to (1.31). The first is when y_3 is always observed. For example, y_3 is education and y_1 is log(wage) and education is always observed no matter the observability of wage. In the second case, y_3 is observed only along with y_1 . An example is when y_1 is log(wage) and y_3 is a measure of nonwage benefits. We do not observe nonwage benefits when we do not observe wage.

From the assumption we made above, we can write

$$y_1 = \beta_1 y_2^* + \beta_2 y_3 + z_1 \gamma + \xi v_2 + e_1 \quad (1.32)$$

where $e_1 \equiv u_1 - E(u_1|v_2)$. And $E(e_1|z, v_2) = 0$. Since e_1 is not correlated with y_2^* , y_2^* is exogenous in (1.32). If v_2 were observed, we could estimate (1.32) by 2SLS on the selected sample using instruments (z, v_2) . As before, we can estimate v_2 when $y_2 > 0$ since π_2 can be consistently estimated by Tobit of y_2 on z .

PROCEDURE 1.2

(i) Obtain $\hat{\pi}_2$ from Tobit of y_2 on z using all observations. Obtain the Tobit

residuals $\hat{v}_{i2} = y_{i2} - z_i \hat{\pi}_2$ for $y_{i2} > 0$.

(ii) Using the selected subsample to estimate the equation

$$y_{i1} = \beta_1 y_{i2} + \beta_2 y_{i3} + z_{i1} \gamma + \xi \hat{v}_{i2} + e_{i1} \quad (1.33)$$

by 2SLS, using instruments (y_2, z_i, \hat{v}_{i2}) .

Allowing for y_3 means that the structural equation can contain standard form of endogeneity apart from sample selection. For example, when y_1 is wage offer and y_2 is hours worked, y_3 might be education. Thus, we allow for self-selection into the model as well as account for the sample selection problem. We discuss self-selection in the next chapter.

VI. Conclusion

This chapter has derived a multi-step approach to estimating a Type III Tobit model where the selection variable appears as an endogenous explanatory variable. The multi-step approach has advantages of being easy to compute and being more robust than MLE. It also provides a simple t-test for sample selection bias at the same time. Computing the asymptotic variance matrices when sample selection is present requires general methods for two-step estimation, as in Newey and McFadden (1994, Handbook of Econometrics.).

When we apply the approach to the Moffitt (1984) data, we find evidence of sample selection bias, contrary to what Moffitt finds. After correcting the bias, the average return of education goes from 7.77% to 5.94%, which is similar to earlier findings. It appears that the average return will be overestimate if we do not account for the sample selection. The coefficient of hours* also changes from .36% to 1.21% before and after we account for endogeneity of hours in the wage offer equation.

As for labor supply, after our three step estimation we found out that the wage elasticity of labor supply is far less when the estimation is conditional on working hours >0 than without, something that has been reported in earlier literature.

We also showed how to relax the assumption that the structural error has a conditional expectation linear in the selection error. The model can be extended to handle other endogenous explanatory variables in the structural equation. The next chapter considers a more general model with endogenous explanatory variables and sample selection.

Table 1.1

Descriptive Statistics for NLS Data Set

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
hours	610	17.1	19.06	0	60
wage	297	2.50	1.06	.44	6.62
nowinc	610	12.50	9.32	0	48.08
marital	610	.80	.40	0	1
age	610	44.37	3.35	35	49
nonwhite	610	.26	.44	0	1
famber	610	4.31	2.06	1	11
chiles6	610	.10	.39	0	4
chigre6	610	1.93	1.74	0	7
educ	610	10.32	2.77	2	18
labforce	610	.41	.84	.002	4.59
indfr1	610	.22	.02	.18	.25
Indfr2	610	.18	.02	.15	.23

Note: hours are weekly hours worked, wage is hourly wage rate, nowinc is the flow of asset income, marital is marital status (=1 if married), nonwhite is a binary variable, which equals one if nonwhite. famber represents the number of family members, chiles6 represents number of kids less than 6 years of age, chigre6 represents number of kids greater than 6 years of age, educ represents years of education, labforce is the size of labor force (in millions), indfr1 is employment fractions in manufacturing in the census region of residence, indfr2 is the employment fractions in government in the census region of residence.

Table 1.2

The Estimation of Different Specifications

	(1) OLS	(2) OLS+ \hat{v}_2	(3) (2)+ hours ^{*2}
Constant	-.9093 (.5789)	-.5153 (.7027)	-.0625 (.7083)
nonwhite	-.1325* (.0485)	-.1626* (.0503)	-.1622* (.0504)
Educ	.0777* (.0078)	.0594* (.0115)	.0594* (.0116)
Labforce	.0923* (.0264)	.1013* (.0266)	.1012* (.0267)
Indfr1	-1.2751* (.6968)	-.3620 (.8149)	-.3534 (.8186)
Indfr2	-4.2511* (1.7563)	-1.9894 (2.0440)	-1.9897 (2.0476)
Age	.0015 (.0061)	.0104 (.0074)	.0103 (.0074)
hours*	.0036* (.0019)	.0121* (.0044)	.0110 (.0090)
hours ^{*2}	-	-	.00002 (.00014)
\hat{v}_2	-	-.0091* (.0044)	-.0093* (.0044)
R-Square	.3612	.3711	.3712

Notes: Dependent variable is log hourly wage.

educ represents years of education.

exper represents experience.

\hat{v}_2 is the Tobit residual.

Standard errors in parentheses.

* : significant at the 10% level.

Table 1.3

Two step and Moffitt's MLE estimation

	(1) Two step procedure with wage as dependent variable	(2) (1)+ hours ^{*2}	(3) OLS	(4) Moffitt's estimation
Constant	.4269 (1.7936)	.3757* (1.8078)	2.04 (1.49)	-1.80 (1.54)
nonwhite	-.1655 (.1283)	-.1676 (.1287)	-.11 (.12)	-.10 (.12)
Educ	.1607* (.0295)	.1606* (.0295)	.20* (.02)	.20* (.02)
Labforce	.1910* (.0679)	.1916* (.0681)	.17* (.06)	.17* (.08)
Indfr1	-1.185 (2.0801)	-1.2251 (2.0894)	-3.02* (1.78)	-3.59* (2.05)
Indfr2	-5.0859 (5.2176)	-5.0847 (5.2261)	-9.51* (4.48)	-10.27* (5.11)
Age	.0285 (.0188)	.0284 (.0189)	.01 (.02)	.01 (.02)
hours [*]	.0243* (.0113)	.0294 (.0231)	.014 (.02)	.053* (.02)
hours ^{*2}		-.00009 (.00035)	-.00011 (.0003)	-.00078* (.0002)
\hat{v}_2	-.0182* (.0111)	-.0181* (.0111)	-	-
R-Square	.3243	.3245	.3182	.1694

Notes: Dependent variable is hourly wage.

educ represents years of education.

exper represents experience.

\hat{v}_2 is the Tobit residual.

Standard errors in parentheses.

* : significant at the 10% level.

Table 1.4

Tobit Estimation for the Labor Supply Equation

(dependent variable is hours)

	Coefficient	Standard error	t value
Constant	49.05	23.44	2.09
lwagehat	19.30	6.30	3.06
nowinc	.009	.17	.05
marital	-5.78	4.14	-1.40
age	-1.20	.49	-2.47
race	5.12	4.05	1.27
famber	-.23	1.73	-.13
chiles6	-4.06	4.64	-.88
chigre6	-.73	1.91	-.38

Number of obs = 610

chi2(8) = 24.05

Prob > chi2 = 0.0022

Log Likelihood = -1694.5467

Obs. summary: 313 left-censored observations at hours≤0

297 uncensored observations

note: lwagehat is estimated log(wage). nowinc is the flow of asset income. Marital is marital status (=1 if married). Race is a binary variable, which equals to one if nonwhite. Famber represents the number of family members. chiles6 represents number of kids less than 6 years of age. chigre6 represents number of kids greater than 6 years of age.

Table 1.5

Two Step Estimation with \hat{v}_2^2 added as an explanatory variable in the structural equation

	coefficient	standard error	t value
constant	-.55	.41	-1.347
race	-.15	.05	-3.09
educ	.06	.01	5.36
labforce	.10	.03	3.15
indfr1	-.36	.82	-.45
indfr2	-2.06	2.05	-1.01
age	.01	.007	1.50
hours	.012	.004	2.93
\hat{v}_2	-.014	.007	-2.03
\hat{v}_2^2	.00009	.0001	.91

Number of obs = 297

R-squared = 0.3718

Notes: Dependent variable is log hourly wage.

educ represents years of education.

exper represents experience.

\hat{v}_2 is the Tobit residual.

Chapter 2

SELECTION CORRECTION WHEN BOTH SELF-SELECTION BIASES AND SAMPLE SELECTION BIASES ARE PRESENT IN A RANDOM COEFFICIENT MODEL

I. Introduction

In chapter one we discussed sample selection correction procedures in models with endogenous variables and a Tobit selection equation. In this chapter, we focus on the situation where there is both self-selection and sample selection. A lot of literature has focused on either corrections for self-selection biases or corrections for sample selection biases before. However, none of them corrects these two sources of biases at the same time.

The problem of self-selection often occurs when evaluating various programs such as job training, welfare, or the returns to education using nonexperimental data. Generally, the issue is that individuals decide whether to participate in a program or to receive some treatment. If the participation decision is related to factors that affect the outcome variable, ignoring the determinates of this decision leads to biased estimates of program impacts. As an example, suppose we are interested in evaluating the benefits of social programs. A common specification is as follows:

$$y_1 = z_1\beta_1 + \alpha y_2 + u_1,$$

where y_1 is an outcome such as earnings, z_1 is a vector of exogenous characteristics, and y_2 is the program variable that can be a dummy or a continuous variable. If we are

interested in the impact of the program on employed persons, it is sufficient to obtain a random sample of people who are currently working. If we are interested in the impact of the program for the population of eligible persons, but we can only obtain a sample of people who work, the problem of sample selection bias arises. In addition, the variable y_2 generally cannot be treated as exogenous if the decision of an individual to attend the program is based on individual self-selection. If we were to treat the variable y as exogenous, the self-selection biases also arise. If the sample we obtain represents the population we are interested, there is no sample selection problem and only the self-selection problem remains. However, if the sample also does not represent the population of interest, we have to address both sample selection and self-selection problems.

An early discussion of the self-selection problem was that of Roy (1951), who discussed the problem of individuals choosing between two professions, hunting and fishing, based on their productivity in each. The observed distribution of incomes of hunters and fishermen was determined by these choices. The result showed that the individuals with better skills go into the profession with higher variance in earnings.

The econometrics of the sample selection problem began with the studies by Gronau (1974), Lewis (1974), and Heckman (1974). Heckman (1976) suggested a two-stage estimation method to address this problem. Subsequently, Lee and Trost (1978) also analyzed the self-selection problem in the case of housing demand. About the same time, Lee (1978) corrected for sample selection biases corrections when estimating the return to union membership, and Willis and Rosen (1979) did the same thing when estimating the rate of return to a college education.

In the foregoing studies, the self-selection problem involved a discrete choice between two alternatives. By contrast, Garen (1984) studied self-selection problem with a continuous choice variable in the case of the returns to schooling where unobserved heterogeneity interacts with a continuous endogenous² explanatory variable. In Garen's application, the effect of schooling on wage may depend on the level of ability and self-selection may cause schooling and ability to be correlated. Garen's model will be discussed in section II.

Angrist and Krueger (1991) also studied the case of the returns to schooling. They explored self-selection problem in the context of how compulsory school attendance laws affect schooling and earnings. They argued that season of birth is related to educational attainment, because of policies regarding the age when children may first start school and compulsory school attendance laws. People born in the beginning of the year start school at an older age and can drop out after completing less schooling than people born near the end of year. They use quarter of birth as instrument for education and apply two-stage least squares (2SLS) to estimate the model. In this chapter, we examine the case that combines both self-selection and sample selection biases with a Tobit selection equation. We also want to compare the different situations under different estimation methods such as OLS, 2SLS, the method proposed by Garen, and the method we propose.

The remainder of the chapter is organized as follows. Section II develops the basic model and examines the implications of different combinations of self-selection and sample selection. Section III develops a model that extends Garen's procedure to handle both self-selection and sample selection. In section IV we use two different data sets from the

² Endogenous variable here means any explanatory variable correlated with unobservables.

labor economics literature to verify the models we proposed in section III. Section V shows consistency of the new procedure we used in section IV. Section VI is the conclusion.

II. A Random Coefficient Model

Garen (1984) considered a model where an endogenous explanatory variable interacts with an unobservable. This kind of model also called a “random coefficient” model. Specifically,

$$y_1 = z_1\beta_1 + \alpha_1 y_2 + \gamma_1 a + \gamma_2 y_2 a + u_1 \quad (2.1)$$

where z_1 represent the exogenous variables and a constant and y_2 is the variable that is correlated with the unobserved variable a . As we will see, the assumptions we impose are most reasonable when y_2 is (roughly) continuous. For identification reasons, we need to assume z (which represent all of the exogenous variables) contains at least one element not in z_1 . The variable u_1 is the structural error, where we assume that $E(u_1 | z, y_2, a) = 0$.

Therefore, (2.1) represents a conditional expectation:

$$E(y_1 | z, y_2, a) = z_1\beta_1 + \alpha_1 y_2 + \gamma_1 a + \gamma_2 y_2 a \quad (2.2)$$

The partial effect of y_2 on $E(y_1 | z, y_2, a)$ depends on a :

$$\frac{\partial E(y_1 | z, y_2, a)}{\partial y_2} = \alpha_1 + \gamma_2 a.$$

Without loss of generality we can assume that $E(a) = 0$. Therefore, the average effect in

the population, sometimes called the average treatment effect, is α_1 . This is the primary parameter of interest.

Garen argued that since y_2 is generally correlated with z , $y_2 a$ is generally correlated with z even if $E(a|z)=0$. So, if we estimate the equation

$$y_1 = z_1 \beta_1 + \alpha_1 y_2 + e_1$$

by 2SLS using instruments z , the estimates of β_1 and α_1 will not be consistent. Garen suggested a correction approach to correct the self-selection biases when there exist the interaction term of y_2 and the unobservable. The reduced form for y_2 is:

$$y_2 = z \theta_2 + v_2. \quad (2.3)$$

Garen assumed $u_1 \sim \text{Normal}(0, \sigma_{u_1}^2)$, $a \sim \text{Normal}(0, \sigma_a^2)$, and $\text{Cov}(u_1, a) = \sigma_{u_1 a}$. Also,

$E(a|z, v_2)$ is independent of z and linear in v_2 :

$$E(a|z, v_2) = E(a|v_2) = \rho_2 v_2. \quad (2.4)$$

Under (2.4), the expected value of (2.2) conditional on (z, v_2) is:

$$\begin{aligned} E(y_1 | z, v_2) &= z_1 \beta_1 + \alpha_1 y_2 + \gamma_1 E(a|z, v_2) + \gamma_2 y_2 E(a|z, v_2) + E(u_1 | z, v_2) \\ &= z_1 \beta_1 + \alpha_1 y_2 + \gamma_1 \rho_2 v_2 + \gamma_2 \rho_2 y_2 v_2 \end{aligned}$$

This suggests a natural two-step procedure proposed by Garen (1984):

1. Regress y_2 on z and save the residuals \hat{v}_2 .
2. Regress y_1 on $z_1, y_2, \hat{v}_2, y_2 \hat{v}_2$.

By the standard theory of generated regressors, the second regression produces consistent estimators of β_1 and α_1 , $\gamma_1 \rho_2$, and $\gamma_2 \rho_2$.

In a recent paper, Wooldridge (1997) showed that Garen's claim about the inconsistency of 2SLS is not correct. Wooldridge showed that 2SLS does produce consistent estimators when we ignore the interaction term in Garen's model under assumptions weaker than Garen imposed. We briefly describe the reasoning as follows:

Rewrite equation (2.2) as

$$y_1 = \delta_1 + z_1\beta_1 + \alpha_1 y_2 + \alpha y_2 + u_1 \quad (2.5)$$

and assume $E(u_1 | z) = 0$. Because equation (2.5) contains the term $\alpha_1 y_2$, we can assume $E(\alpha) = 0$ without loss the generality. For y_2 , we can write a linear equation in terms of all exogenous variables as in (2.3):

$$y_2 = \delta_2 + z\theta_2 + v_2 = \delta_2 + z_1\theta_{21} + z_2\theta_{22} + v_2 \quad (2.6)$$

If v_2 is correlated with α , then y_2 is endogenous and the expected error term will not be zero, which is:

$$E(\alpha y_2 + u_1) = E(\alpha y_2) = E(\alpha v_2) \neq 0 \quad (2.7)$$

Because of this, Garen argued that the 2SLS estimator α_1 would be inconsistent.

However, 2SLS only causes the intercept term δ_1 to be inconsistently estimated and the other parameters will be consistently estimated. To see this, we need to first assume $E(u_1 | z) = 0$ as we already did and we need to assume v_2 satisfies the zero mean conditional and homoskedasticity assumptions from standard linear regression analysis.

That is, $E(v_2 | z) = 0$ and $E(v_2^2 | z) = \sigma_2^2$. We also need to assume a relationship between α and v_2 as before, $E(\alpha | z, v_2) = E(\alpha | v_2) = \rho_2 v_2$. This means α is conditional mean independent of z given v_2 and $E(\alpha | v_2)$ is linear. We need the standard identification condition that $\theta_{22} \neq 0$ in (2.6). Now, since

$$\begin{aligned} E(\alpha y_2 | z) &= E[E(\alpha y_2 | z, v_2) | z] = E[y_2 E(\alpha | z, v_2) | z] \\ &= E(\rho_1 y_2 v_2 | z) = \rho_1 E(v_2^2 | z) = \rho_1 \sigma_2^2 = E(\alpha y_2), \end{aligned}$$

we have

$$y_1 = (\delta_1 + \rho_1 \sigma_2^2) + z_1 \beta_1 + \alpha_1 y_2 + r_1, \quad (2.8)$$

where $r_1 = [\alpha y_2 - E(\alpha y_2 | z)] + u_1$, so $E(r_1 | z) = 0$.

Applying 2SLS to (2.8) using a random sample of size n , where z_2 is the vector of instruments for y_2 , is consistent and \sqrt{n} -asymptotically normal for β_1 and α_1 under standard finite moment conditions. For more details refer to the paper of Wooldridge.

There is one thing that should be noticed: although Garen's claim about the inconsistency of 2SLS is incorrect, that does not mean the estimation method he proposed will produce an inconsistent estimate. Both 2SLS and Garen's model should work under similar assumptions.

III. Garen's Model with Sample Selection

Garen (1984) assumed that a random sample from the underlying population is available. This may not be the case in his example, which is to estimate the average return to education in a wage offer equation. As we discussed in chapter one, using only observations on people who are working can cause a sample selection bias.

In this section we study estimation of Garen's model with a Tobit sample selection mechanism. It can be written as:

$$y_1 = z_1 \beta_1 + \alpha_1 y_2 + \delta_1 y_2 \alpha_1 + u_1 \quad (2.9)$$

$$y_2 = z\pi_2 + v_2 = z_1\pi_{21} + z_2\pi_{22} + v_2 \quad (2.10)$$

$$y_3 = \max(0, z\pi_3 + v_3), \quad (2.11)$$

where y_1 is observed when $y_3 > 0$, z represent the exogenous variables and a constant, and y_2 is the endogenous variable that is possibly correlated with the unobserved variable a . The assumptions below are most suited to the case when y_2 is continuous. For identification reason, we need to assume z (which represent all of the exogenous variables) contains at least one element not in z_1 .

We have added a Tobit selection equation to Garen's model. An example of this is when y_1 is the log of the hourly wage, y_2 is years of education and y_3 is weekly or annual hours of works. More details about this example will be given in section III.

Assumptions 2.1

- (i) (z, y_2, y_3) is always observed in the population but y_1 is observed only when $y_3 > 0$.
- (ii) (u_1, v_2, v_3, a_1) is zero-mean independent of z .
- (iii) $v_3 \sim \text{Normal}(0, \tau_3^2)$.
- (iv) $E(a_1 | z, v_2, v_3) = \rho_1 v_2 + \rho_2 v_3$.
- (v) $E(u_1 | z, v_2, v_3) = \zeta_1 v_2 + \zeta_2 v_3$.
- (vi) The rank condition $\pi_{22} \neq 0$.

Assumption (i) says that y_1 is the only variable unobserved due to a possible sample selection problem. Assumption (ii) and (iii) are fairly standard in this contexts, but they are restrictive. Assumption (ii) means that the disturbance term in the linear reduced form for y_2 (which is v_2) is independent of z , this restricts the kinds of endogenous variables we

can allow. If y_2 is a binary or discrete variable, then the assumption is not reasonable. This could be relaxed somewhat, but linearity is still needed.

Assumption (iii) allows u_1, v_2 , and v_3 to be arbitrarily correlated, so that endogeneity and sample selection bias can both be present. Assumption (iv) implies that the conditional expectation involving unobservable is linear. Assumption (v) relaxes the usual joint normality of (u_1, v_2, v_3) by just requiring linearity of a conditional expectation. Assumption (vi) is the standard identification condition: We need a good instrumental variable for y_2 .

The model is a Type III Tobit model in Amemiya's (1985) taxonomy but with an endogenous explanatory variable that interacts with an unobservable.

To derive an estimating equation, write

$$\begin{aligned} E(y_1 | z, v_2, v_3) &= z_1 \beta_1 + \alpha_1 y_2 + \delta_1 y_2 E(\alpha_1 | z, v_2, v_3) + E(u_1 | z, v_2, v_3) \\ &= z_1 \beta_1 + \alpha_1 y_2 + \delta_1 y_2 (\rho_1 v_2 + \rho_2 v_3) + \zeta_1 v_2 + \zeta_2 v_3 \end{aligned}$$

by assumptions (ii), (iv), and (v). So we have

$$E(y_1 | z, v_2, v_3) = z_1 \beta_1 + \alpha_1 y_2 + \kappa_1 y_2 v_2 + \kappa_2 y_2 v_3 + \zeta_1 v_2 + \zeta_2 v_3$$

where $\kappa_1 \equiv \delta_1 \rho_1$ and $\kappa_2 \equiv \delta_1 \rho_2$. From the discussion in section II of chapter one, we can select the sample on the basis of (z, v_3) .

Since v_2 and v_3 are not observed, we can estimate v_2 by running OLS on equation (2.9) and estimate v_3 when $y_3 > 0$ since π_3 can be consistently estimated by Tobit of y_3 on z . The procedure is:

PROCEDURE 2.1

- (i) Run OLS of y_2 on z to get $\hat{\pi}_2$. Obtain the OLS residuals $\hat{v}_2 = y_2 - z\hat{\pi}_2$.
- (ii) Obtain $\hat{\pi}_3$ from Tobit of y_3 on z using all observations. Obtain the Tobit residuals $\hat{v}_3 = y_3 - z\hat{\pi}_3$ for $y_3 > 0$.
- (iii) Use OLS on the selected subsample for which we observe y_1 to estimate the equation

$$y_1 = z_1\beta_1 + \alpha_1 y_2 + \kappa_1 y_2 \hat{v}_2 + \kappa_2 y_2 \hat{v}_3 + \zeta_1 \hat{v}_2 + \zeta_2 \hat{v}_3 + \text{error term}.$$

We can test and correct the potential self-selection and sample selection biases at the same time by procedure 2.1. An F-test of the joint significance of $y_2 \hat{v}_2$, $y_2 \hat{v}_3$, \hat{v}_2 , and \hat{v}_3 tests for endogenous of y_2 or sample selection. If either $\hat{\kappa}_1$ or $\hat{\zeta}_1$ is significant, there is a self-selection problem. If either $\hat{\kappa}_2$ or $\hat{\zeta}_2$ is significant, there is a sample selection problem. If we use individual t-statistics or F-statistics for a subset of coefficients, these should be adjusted for the generated regressors problem. If the coefficients on these variables are small, the adjustment will not make much of difference.

Notice that the exogenous explanatory variables z in (2.10) and (2.11) can be the same: we do not need an exclusion restriction in (2.10) in order for the procedure to work well. The residuals \hat{v}_3 have separate variation from $z\hat{\pi}_3$ because of the variation in y_3 .

IV. Empirical Examples

As in Chapter 1, we use an example from labor economics as the basis for the empirical applications to follow.

A wage offer labor supply system is given by

$$\log(\text{wage}^0) = z_1 \beta_1 + \alpha_1 \text{educ} + \delta_1 \text{educ} \cdot \alpha_1 + \alpha_1 + u_1 \quad (2.12)$$

$$\text{educ} = z\pi_2 + v_2 \quad (2.13)$$

$$\text{hours} = \max(0, z\pi_3 + v_3) \quad (2.14)$$

where wage^0 is the wage offer observed only when $\text{hours} > 0$, educ is years of education, α_1 is unobserved ability. Education is correlated with unobserved ability. There exists an interaction term $\text{educ} \cdot \alpha_1$ for education and ability. We assume that hours , z , and educ are always observed. The assumptions are the same as in section II. We are interested in estimating the returns to education in this model, and we do so using different estimation methods including OLS, 2SLS, the method suggested by Garen, and the method proposed in Section III.

We first apply these methods to the data set used by Mroz (1987) in his study of the sensitivity of female labor supply to various assumptions. The data come from the University of Michigan Panel Study of Income Dynamics (hereafter PSID) for the year 1975. The sample consists of 753 married women, 428 of whom report positive hours worked during 1975. The exogenous variables include actual labor market experience (exper), exper^2 , family income other than that earned by the woman – in thousands – (nwfeinc), number of kids less than six (kidslt6), number of kids between six and eighteen (kidsge6), and age. We make exclusion restrictions in the wage equation, which includes only educ , exper , and exper^2 as explanatory variables. Hours here are measured annually. The descriptive statistics are summarized in Table 2.1.

We summarize the estimation results for the PSID data set in Table 2.2. In the OLS estimation, we estimate the model using only the women who work, which is ignoring the potential sample selection biases and self-selection biases. The result shows in column (1) of Table 2.2.

The OLS estimate return of the education using the PSID data set is about 10.75%. However, in our model educ is endogenous and possibly correlated with unobserved ability. It is possible OLS estimation is biased. So we want to see if there exists self-selection biases and sample selection biases problems by checking the other versions of estimation.

To see if there self-selection bias exists, we first run a regression which only include \hat{v}_2 as the additional regressor. The result shows in column (2) of Table 2.2.

The t-test for the estimated coefficient of \hat{v}_2 is marginally significant, suggesting possible self-selection bias. To correct the possible self-selection bias, we use 2SLS and Garen's method to make comparisons. The results of 2SLS and Garen's estimation is in column (3) and (4) of Table 2.2.

The 2SLS estimate of returns to education is 8.04%. We use parents education and husband education as the instrument variables for educ. The parents education levels are widely believed good instruments for educ although the husband education level is more suspectable to be a good instrument. According to section II, the estimation is consistent if there is only self-selection biases present. Since $\hat{v}_2 = educ - z\hat{\pi}_2$, it can be proved that the OLS estimators of the returns to education are identical to the 2SLS estimators when we add \hat{v}_2 as an additional regressor in a normal OLS. Although the

parameters estimate of 2SLS and the one we did in column (2) are the same, the procedure of adding \hat{v}_2 as an additional regressor in a normal OLS is useful as a test to see if there exists self-selection biases.

We also perform the estimation method proposed by Garen to compare with 2SLS. The results are shown in column (4). In column (4), the return to education is 7.64%. If we compare the estimated coefficients for the returns to education in column (3) and (4), we can see that there is only 0.4 percentage point difference.

To see if there exists sample selection biases in addition to self-selection biases, we suggest two models. One includes both corrections for self-selection and sample selection biases but linear model which we will call it procedure 2.2 afterwards. More details for procedure 2.2 are given in section V. The other one is our new method developed above which include the linear terms \hat{v}_2, \hat{v}_3 and the nonlinear terms $\text{educ} \cdot \hat{v}_2, \text{educ} \cdot \hat{v}_3$. The results are shown in column (5) and (6).

The return to education in column (5) is 7.65%, which is similar to the results in 2SLS and Garen's model. This is reasonable since the t value for the estimator of coefficient of \hat{v}_3 is -1.351 , which indicates the sample selection biases is not a server problem here.

In column (6), the returns to education is 7.98%. The result is very close to the estimation by 2SLS, Garen's model, and the linear combined model we proposed above. From the t value of each estimate coefficient, we have the same conclusion as before that there only exists a self-selection problem. We also make some joint tests for self-selection biases and sample selection biases. The results show in Table 2.3. It also shows that there

is strong evidence that there exist self-selection biases but no signs of sample selection biases.

We notice there is about a 20% difference for the estimation of returns to education between OLS and other estimation methods. We can conclude that the OLS estimator of α_1 is inconsistent. Also, the 2SLS version of $\hat{\alpha}_1$ is .0804, which is not much difference from the Garen's version of $\hat{\alpha}_1$ (about 0.076). Sample selection does not appear to be a problem in this case.

We now apply the same procedure to another, larger data set, which was compiled from the May 1991 Current Population Survey (CPS) by Daniel Hamermesh. The data set contains 5634 married women, 3286 of whom report working positive hours during the week. The hourly wage is weekly earnings divided by weekly hours for women who worked positive hours; the women did not work do not have data for hourly wage. The hours here is measured weekly rather than annually as in PSID data set. Also, experience is potential rather than actual experience and the variables kidslt6 and kidsge6 are binary indicators. The descriptive statistic are shown in Table 2.4.

As in Table 2.2, here we perform all the estimation methods for CPS data set in Table 5. OLS results are shown in column (1).

In column (1), the estimated return to an additional year of education is 9.90%. To see if there exist self-selection biases, we add \hat{v}_2 as an additional explanatory variable for OLS. The results are shown in column (2).

From the t statistics for the estimated coefficient of \hat{v}_2 , it is clear that there does exist self-selection biases and need to be corrected. We find the returns to education is

13.05% in this case and the t statistic is significant at the 10% level. Although it is widely believed that estimate of the return to education should go down after correction of the self-selection, there is growing research supporting our empirical results of the CPS's data set that OLS estimate is biased downward. Angrist and Krueger (1991) in their research for returns to schooling reach the same result that OLS estimates are biased downward than instrumental variables estimates.³ David Card (1993) also reach the same result that instrumental variable estimates exceed OLS estimate and suggest that the result may come from measurement error in schooling.

The 2SLS results are shown in column (3). As we said, the estimators of 2SLS are algebraically the same as we add \hat{v}_2 as additional regressor in OLS. So the estimated coefficients is just the same as the result in previous column. Since the CPS data set we have does not have other good instruments for educ, we can only use husband education as the only instrument variable for educ. If there is no sample selection biases present, the estimate should be consistent. For comparison reason, the estimation of Garen's method is shown in column (4).

The return to education in Garen's model is 13.07%, which is about the same as in 2SLS version. This support the claim made by Wooldridge that 2SLS estimator is, under certain assumptions, consistent even when we allow the endogenous explanatory variable to interact with unobservables. When the data set is bigger, we observe that the estimators of 2SLS and Garen's model is closer than in PSID data set. We also have to check if the sample selection biases is a problem. To see that, we also run the two methods we

³ Other researches like Ashenfelter and Krueger (1992), Kane and Rouse (1993) and Butcher and Case (1993). All the above studies report instrumental variables estimates of the return to schooling that exceed the ordinary least-squares estimate.

proposed as we did in PSID data set case. The estimation result, which makes corrections for self-selection and sample selection biases but linear model, is in column (5).

From the t-test of the estimated coefficient for \hat{v}_3 , it's clear that there exist sample selection biases problem. The t value is 2.503, which is pretty significant. The returns to education rise a little to 13.09% from 2SLS. The estimation including the interactive terms as additional regressors shows in column (6).

In column (6), the return to education we estimated is 13.21%, which is pretty close to the estimation method that only include \hat{v}_2 and \hat{v}_3 as additional regressors. We notice that from t-test, both estimate coefficients of \hat{v}_3 and $\text{educ} \cdot \hat{v}_3$ are not significant. It indicates there is multicollinearity present.

We do the F-test to jointly test the four terms including \hat{v}_2 , \hat{v}_3 , $\text{educ} \cdot \hat{v}_2$, and $\text{educ} \cdot \hat{v}_3$. The result shows in Table 2.6. We can see that the joint test for estimate coefficients of $\text{educ} \cdot \hat{v}_3$ and \hat{v}_3 is marginally significant at the 10% level. This supports our claim that the sample selection biases do exist in this case.

It is important to note that the estimated return to education by OLS is about the same (10.75% for PSID data set and 9.9% for CPS data set) before the self-selection and sample selection problems are being controlled for. After correcting self-selection and sample selection biases, the return to education decreases to 7.98% in PSID data set and increases to 13.21% in CPS's data set. From the estimation results in Table 2.2 and 2.5, this OLS bias of return to education mainly comes from people's self-selection.

To make the exact comparison with CPS data set, we redo the procedure 2.1 for PSID data set and remove parents' education as instruments for education. This leaves

husband education as the only instrument for education just like in CPS data set. The results are shown in Table 2.7.

In Table 2.7, both self-selection and sample selection don't appear to be problems. The result is different from Table 2.2 where we use parents' education and husband's education as instruments of education. This result indicates that husband's education may not be a good instrument for education. Compare with the results in Table 2.5 and 2.7, the situation that OLS estimations are close but procedure 2.1 estimations are very different keeps the same.

V. The Consistency of Procedure 2.2

When there is only a self-selection problem, both 2SLS and Garen's estimation method can produce consistent estimates for the return to education under the assumptions discussed in section III. We also mentioned in section III that the OLS estimators of the returns to education are identical to the 2SLS estimators when we add \hat{v}_2 as an additional regressor in a normal OLS. This means both OLS y_1 on $z_1, y_2, \hat{v}_2, y_2 \hat{v}_2$ and OLS y_1 on z_1, y_2, \hat{v}_2 can get consistent estimators of α_1 (the returns to education).

In procedure 2.1 we regress y_1 on $z_1, y_2, y_2 \hat{v}_2, y_2 \hat{v}_3, \hat{v}_2, \hat{v}_3$. It is natural for us to suspect that OLS of y_1 on $z_1, y_2, \hat{v}_2, \hat{v}_3$ can actually produce consistent estimator of α_1 . Compare the results of Procedure 2.2 and procedure 2.1 in Table 2.2 and 2.5, we also find that $\hat{\alpha}_1$ in both methods are very close.

Our conjecture turns out to be correct. We show it in the following.

As in section II, our model is:

$$y_1 = c_1 + z_1\beta_1 + \alpha_1 y_2 + \delta_1 y_2 \alpha_1 + u_1 \quad (2.9)'$$

$$y_2 = z\pi_2 + v_2 = z_1\pi_{21} + z_2\pi_{22} + v_2 \quad (2.10)$$

$$y_3 = \max(0, z\pi_3 + v_3) \quad (2.11)$$

where c_1 in (2.9)' is the intercept. We also need certain assumptions:

Assumptions 2.2

- (i) $E(u_1 | z) = 0$
- (ii) $E(v_2 | z) = 0$, $E(v_2^2 | z) = \sigma_2^2$
- (iii) $E(v_3 | z) = 0$, $\text{Cov}(v_2, v_3) = 0$
- (iv) $E(\alpha_1 | z, v_2, v_3) = E(\alpha_1 | v_2, v_3) = \rho_1 v_2 + \rho_2 v_3$
- (v) $E(u_1 | z, v_2, v_3) = \zeta_1 v_2 + \zeta_2 v_3$
- (vi) $\pi_{22} \neq 0$ in (2.10).

Assumption (i) is an exogenous assumption for z . Assumption (ii) implies that v_2 satisfies the zero conditional mean and homoskedasticity assumption from standard linear regression analysis. Assumption (iii) means that v_3 satisfies the zero conditional mean and is uncorrelated with v_2 . Assumption (iv) allows y_2 and y_3 to be correlated with α_1 . Assumption (v) allows y_2 and y_3 to be correlated with u_1 . Assumption (vi) is the standard rank condition for identification.

We can show that by only adding OLS residuals of equation (2.10) (denoted as \hat{v}_2) and Tobit residuals of equation (2.11) (denoted as \hat{v}_3) as additional explanatory

variables in the structure equation (2.9)', β_1 and α_1 can be consistently estimated. First we need to show that $E(a_1 y_2 | z, v_2, v_3) = E(a_1 y_2)$. Under (2.10), (ii), (iii), (iv), and the law of iterated expectations,

$$\begin{aligned}
E(a_1 y_2 | z, v_2, v_3) &= E[E(a_1 y_2 | z, v_2, v_3) | z, v_2, v_3] = E[y_2 E(a_1 | z, v_2, v_3) | z, v_2, v_3] \\
&= E[z \pi_2 (\rho_1 v_2 + \rho_2 v_3) | z, v_2, v_3] + E[v_2 (\rho_1 v_2 + \rho_2 v_3) | z, v_2, v_3] \\
&= E(z \pi_2 \rho_1 v_2 | z, v_2, v_3) + E(z \pi_2 \rho_2 v_3 | z, v_2, v_3) \\
&\quad + E(\rho_1 v_2 v_2 | z, v_2, v_3) + E(\rho_2 v_2 v_3 | z, v_2, v_3) \\
&= \rho_1 E(v_2^2 | z, v_2, v_3) = \rho_1 \sigma_2^2 = E(a_1 y_2)
\end{aligned}$$

After we have the proof above, we can write

$$y_1 = (c_1 + \delta_1 \rho_1 \sigma_2^2) + z_1 \beta_1 + \alpha_1 y_2 + r_1, \quad (2.12)$$

where $r_1 = [\delta_1 a_1 y_2 - \delta_1 E(a_1 y_2 | z, v_2, v_3)] + u_1$, and so

$$\begin{aligned}
E(r_1 | z, v_2, v_3) &= E(u_1 | z, v_2, v_3) \\
&= \zeta_1 v_2 + \zeta_2 v_3
\end{aligned} \quad (2.13)$$

by assumption (vi).

It follows that we can write

$$E(y_1 | z, v_2, v_3) = (c_1 + \delta_1 \rho_1 \sigma_2^2) + z_1 \beta_1 + \alpha_1 y_2 + \zeta_1 v_2 + \zeta_2 v_3, \quad (2.14)$$

Now let s be a binary indicator such that $s = 1[y_3 > 0]$ which is a function of (z, v_3) , we

have

$$E(y_1 | z, v_2, v_3, s = 1) = (c_1 + \delta_1 \rho_1 \sigma_2^2) + z_1 \beta_1 + \alpha_1 y_2 + \zeta_1 v_2 + \zeta_2 v_3. \quad (2.15)$$

Equation (2.14) leads to the following three-step procedure.

PROCEDURE 2.2

- (i) Run OLS of y_2 on z to get $\hat{\pi}_2$. Obtain the OLS residuals $\hat{v}_2 = y_2 - z\hat{\pi}_2$
- (ii) Obtain $\hat{\pi}_3$ from Tobit of y_3 on z using all observations. Obtain the Tobit residuals $\hat{v}_3 = y_3 - z\hat{\pi}_3$ for $y_3 > 0$.
- (iii) Estimate $\beta_1, \alpha_1, \zeta_1, \zeta_2$ from OLS regression

$$y_1 \text{ on } z_1, y_2, \hat{v}_2, \hat{v}_3$$

using the selected subsample.

It turns out the Procedure 2.1 is also unnecessary just like Garen's method in self-selection case for us to consistently estimate β_1 and α_1 . Both Procedure 2.1 and Procedure 2.2 can estimate the returns to education consistently in section IV's application.

VI. Conclusion

This chapter shows how to test and to correct for self-selection and sample selection biases at the same time, especially in circumstance when the explanatory variable, such as education, interacted with unobservable.

The procedure developed in this paper to test and correct potential self-selection and sample selection biases is the following:

1. If self-selection and sample selection questions exist at the same time, then use the two methods we provide in previous sections (Procedure 2.1 and 2.2) to test and correct the biases.
2. If only self-selection appears to be a problem, add \hat{v}_2 as an additional regressor and run OLS. If \hat{v}_2 is significant and self-selection biases are present, use Garen's method or 2SLS.

It should be noted that Procedure 2.1 and 2.2 produce very similar results in both data sets, especially in the larger CPS data set. This supports our claim that the estimation in Procedure 2.2 is consistent even when there exists an interactive term as a regressor.

It is needed to mention that in this chapter we have avoided full parametric assumptions in deriving the sample selection and self-selection procedures. Nevertheless, we have assumed that the selection variable follows a standard Tobit equation and that the conditional expectation for unobservables is linear. In addition, we have assumed independence between disturbances and exogenous variables instead of weaker and preferable zero conditional mean assumptions. For further researches, it is possible to relax some of the parametric assumptions to make the estimation more robust. Powell (1984) extends least absolute deviations (LAD) estimation to the regression with non-negativity of the dependent variable, and gives conditions under which this estimator is consistent and asymptotically normal. The LAD minimize the sum of absolute deviations of the selection variable from its median function under the assumptions that the error term is continuously distributed with median zero and the median is unique. The restriction for the normality of the error term can be relaxed in this setting. It will be interesting to explore and apply LAD to our case.

The empirical results for the PSID data set shows that there is evidence of self-selection biases but not strong evidence of sample selection biases. The simple OLS regression using only women who work cannot produce consistent estimates and the estimate of return to education is biased upward. By contrast, 2SLS can provide consistent estimates when sample selection bias is not a problem.

In the CPS data set, both self-selection biases and sample selection biases exist. The method we suggested should be used to correct self-selection biases along with sample selection biases. The OLS estimate for returns to education is biased downward. This indicates there is no certain direction for OLS biases.

Also, the estimated return to education by OLS is about the same (10.75% for PSID data set and 9.9% for CPS data set) before the self-selection and sample selection problems are being controlled for. After correcting self-selection and sample selection biases, the return to education decreases to 7.98% in PSID data set and increases to 13.21% in CPS's data set. From the estimation results in Table 2.2 and 2.5, this OLS bias of return to education mainly comes from people's self-selection. Although it is widely believed that estimate of the return to education should go down after correction of the sample selection along with self-selection, there are growing researches supporting our empirical results of CPS's data set that OLS estimate is biased downward. Angrist and Krueger (1991) in their research for returns to schooling reach the same result that OLS estimates are biased downward than instrumental variables estimates.⁴ David Card (1993)

⁴ Other researches like Ashenfelter and Krueger (1992), Kane and Rouse (1993) and Butcher and Case(1993). All the above studies report instrumental variables estimates of the return to schooling that exceed the ordinary least-squares estimate.

also reach the same result that instrumental variable estimates exceed OLS estimate and suggest that the result may come from measurement error in schooling.

The method we develop in this chapter is easy to implement in conventional empirical studies. It can correct self-selection and sample selection biases at the same time. It is hoped that researchers in many fields might benefit from it when analyzing empirical questions.

Table 2.1

The Descriptive Statistics for PSID Data Set

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
hours	753	740.58	871.31	0	4950
kidslt6	753	.238	.5240	0	3
kidsge6	753	1.35	1.32	0	8
age	753	42.54	8.07	30	60
educ	753	12.29	2.28	5	17
wage	753	2.37	3.24	0	25
husage	753	45.12	8.06	30	60
huseduc	753	12.49	3.02	3	17
motheduc	753	9.25	3.37	0	17
fatheduc	753	8.81	3.57	0	17
exper	753	10.63	8.07	0	45
nwifeinc	753	20.13	11.63	-.03	96
lwage	428	1.19	.72	-2.05	3.22
expersq	753	178.04	249.63	0	2025

Note: hours are annual hours worked. kidslt6 represents number of kids less than 6 years of age. kidsge6 represents number of kids greater than 6 years of age. Wage is hourly wage rate. educ represents years of education. husage represents age of husband. huseduc is years of education for husband. motheduc is years of education for mother. fatheduc is years of education for father. exper is experience. nwifeinc is non-wife income. lwage is log wage. expersq is experience²

Table 2.2
The Estimation of Different Specifications
(PSID Data Set)

	(1) OLS	(2) OLS+ \hat{v}_2	(3) 2SLS	(4) Garen	(5) Procedure 2.2	(6) Procedure 2.1
Constant	-.5220* (.1986)	-.1869 (.2836)	-.1869 (.2854)	-.1665 (.2829)	-.0790 (.2888)	-.1572 (.3256)
educ	.1075* (.0141)	.0804* (.0216)	.0804* (.0217)	.0764* (.0217)	.0765* (.0216)	.0798* (.0245)
exper	.0416* (.0132)	.0431* (.0132)	.0431* (.0133)	.0429* (.0131)	.0388* (.0137)	.0388* (.0137)
exper ²	-.0008* (.0004)	-.0009* (.0004)	-.0009* (.0004)	-.0009* (.0004)	-.0008* (.0004)	-.0008* (.0004)
educ · \hat{v}_2	-	-	-	.0107* (.0056)	-	.0108* (.0056)
educ · \hat{v}_3	-	-	-	-	-	-.00001 (.00002)
\hat{v}_2	-	.0472* (.0286)	-	-.0820 (.0736)	.0519* (.0285)	-.0795 (.0732)
\hat{v}_3	-	-	-	-	-.00006 (.00004)	.0001 (.0002)
R-Square	.1568	.1622	.1495	.1694	.1661	.1746

Notes: Dependent variable is log hourly wage.

educ represents years of education.

exper represents experience.

$$\hat{v}_2 = \text{educ} - z\hat{\pi}_2$$

\hat{v}_3 is the Tobit residual.

Standard errors in parentheses.

* : significant at the 10% level.

General I represents correction for both self-selection and sample selection biases

in linear model, without interactive between years of education and ability.

Table 2.3

Joint tests for self-selection and Sample selection biases
(PSID data set)

$H_0 :$	F value	Prob > F
$\text{educ} \cdot \hat{v}_2 = 0$ $\text{educ} \cdot \hat{v}_3 = 0$ $\hat{v}_2 = 0$ $\hat{v}_3 = 0$	2.27	0.0612
$\text{educ} \cdot \hat{v}_2 = 0$ $\hat{v}_2 = 0$	3.54	0.0299
$\text{educ} \cdot \hat{v}_3 = 0$ $\hat{v}_3 = 0$	1.32	0.2688
$\text{educ} \cdot \hat{v}_2 = 0$ $\text{educ} \cdot \hat{v}_3 = 0$	2.18	0.1145

Notes: educ denotes years of education.

$$\hat{v}_2 = \text{educ} - z\hat{\pi}_2$$

\hat{v}_3 is the Tobit residual

Table 2.4**Descriptive Statistics for Hamermash Data Set**

Variable	Observations	Mean	Standard Deviation	Minimum	Maximum
hours	5634	20.72	19.40	0	120
kidslt6	5634	.279	.4487	0	1
kidsge6	5634	.308	.4615	0	1
age	5634	39.43	9.99	18	59
educ	5634	12.98	2.62	0	18
wage	3286	10.37	7.03	.03	200
husage	5634	42.45	11.23	19	86
huseduc	5634	13.15	2.98	0	18
exper	5634	20.44	10.45	0	52
nwifeinc	5634	30269.23	27211.58	0	112500
lwage	3286	2.20	.525	-3.40	5.30
expersq	5634	527.04	468.29	0	2704

Note: hours are weekly hours worked. kidslt6 represents dummy variable if there are kids less than 6 years of age. kidsge6 represents dummy variable if there are kids greater than 6 years of age. Wage is hourly wage rate. educ represents years of education. husage represents age of husband. huseduc is years of education for husband. exper is experience. nwifeinc is non-wife annual income. lwage is log wage. expersq is experience².

Table 2.5

**The Estimation of Different Specifications
(CPS Data Set)**

	(1) OLS	(2) OLS+ \hat{v}_2	(3) 2SLS	(4) Garen	(5) Procedure 2.2	(6) Procedure 2.1
Constant	.6504* (.0587)	.2044* (.0950)	.2044* (.0967)	.1895* (.0950)	.1901* (.0921)	.1629 (.1143)
educ	.0990* (.0035)	.1305* (.0063)	.1305* (.0065)	.1307* (.0063)	.1309* (.0061)	.1321* (.0078)
exper	.0198* (.0033)	.0195* (.0033)	.0195* (.0033)	.0195* (.0033)	.0186* (.0033)	.0187* (.0033)
exper ²	-.0003* (.00008)	-.0003* (.00008)	-.0003* (.00008)	-.0003* (.00008)	-.0003* (.00008)	-.0003* (.00008)
educ · \hat{v}_2	-	-	-	.0036* (.0009)	-	.0031* (.0011)
educ · \hat{v}_3	-	-	-	-	-	-.00008 (.00019)
\hat{v}_2	-	-.0453* (.0076)	-	-.0912* (.0143)	.0426* (.0073)	-.0831* (.0163)
\hat{v}_3	-	-	-	-	-.0016* (.0007)	.0026 (.0027)
R-Square	.2047	.2132	.1854	.2166	.2147	.2177

Notes: Dependent variable is log hourly wage.

educ represents years of education.

exper represents experience.

$$\hat{v}_2 = \text{educ} - z\hat{\pi}_2$$

\hat{v}_3 is the Tobit residual.

Standard errors in parentheses.

* : significant at the 10% level.

General I represents correction for both self-selection and sample selection biases in linear model, without interactive between years of education and ability.

Table 2.6

Joint tests for self-selection and Sample selection biases
(CPS data set)

$H_0 :$	F value	Prob > F
$\text{educ} \cdot \hat{v}_2 = 0$ $\text{educ} \cdot \hat{v}_3 = 0$ $\hat{v}_2 = 0$ $\hat{v}_3 = 0$	13.64	0.0000
$\text{educ} \cdot \hat{v}_2 = 0$ $\hat{v}_2 = 0$	21.42	0.0000
$\text{educ} \cdot \hat{v}_3 = 0$ $\hat{v}_3 = 0$	2.78	0.0622
$\text{educ} \cdot \hat{v}_2 = 0$ $\text{educ} \cdot \hat{v}_3 = 0$	6.34	0.0018

Notes: educ denotes years of education.

$$\hat{v}_2 = \text{educ} - z\hat{\pi}_2$$

\hat{v}_3 is the Tobit residual

Table 2.7

The Estimation of Different Specifications
(PSID Data Set without parents' education as instruments)

	(1) OLS	(2) OLS+ \hat{v}_2	(3) 2SLS	(4) Garen	(5) General I	(6) Procedure 2.1
Constant	-.5220* (.1986)	-.2981 (.3094)	-.2981 (.3189)	-.2625 (.3090)	-.2106 (.3027)	-.2713 (.3395)
educ	.1075* (.0141)	.0894* (.0238)	.0894* (.0231)	.0837* (.0239)	.0880* (.0232)	.0895* (.0261)
exper	.0416* (.0132)	.0426* (.0132)	.0426* (.0132)	.0427* (.0132)	.0375* (.0138)	.0377* (.0137)
exper ²	-.0008* (.0004)	-.0008* (.0004)	-.0008* (.0004)	-.0008* (.0004)	-.0007* (.0004)	-.0007* (.0004)
educ · \hat{v}_2	-	-	-	.0104* (.0056)	-	.0106* (.0055)
educ · \hat{v}_3	-	-	-	-	-	-.00001 (.00002)
\hat{v}_2	-	.0280 (.0296)	-	-.0977 (.0725)	.0293 (.0288)	-.0990 (.0722)
\hat{v}_3	-	-	-	-	-.00006 (.00004)	.0001 (.0002)
R-Square	.1568	.1586	.1536	.1657	.1625	.1712

Notes: Dependent variable is log hourly wage.

educ represents years of education.

exper represents experience.

$$\hat{v}_2 = \text{educ} - z\hat{\pi}_2$$

\hat{v}_3 is the Tobit residual.

Standard errors in parentheses.

* : significant at the 10% level.

General I represents correction for both self-selection and sample selection biases in linear model, without interactive between years of education and ability.

Chapter 3

MODEL SELECTION TESTS FOR TWO-PART MODELS

I. Introduction

In econometrics much of the literature concerned with estimation and inference from a sample of economic data deals with a situation when the statistical model is correctly specified. Consequently, in much of econometric practice, it is customary to assume that the parameterized linear statistical model used for purposes of inference is consistent with the sampling process from which the sample observations were generated. In this ideal case, statistical theory provides techniques for obtaining point and interval estimators of the population parameters and for hypothesis testing. In practice, however, the possibilities for model misspecification are numerous and false statistical models are most likely the rule rather than the exception.

Choosing between two competing models is relatively straightforward when one model is nested within another. Assuming that standard regularity conditions hold, Wald, Lagrange multiplier, and likelihood ratio tests (or quasi-LR tests) can be used. See, for example, Fin and Schmidt (1984). See also Chow (1983, Ch. 9).

When competing models are not strictly nested, which means one model cannot be obtained from the other by imposing appropriate restrictions or as a limiting suitable approximation, statistically choosing between them is more difficult. Two very different approaches have been proposed.

One is based on the work of Cox (1961, 1962) who derived specification tests that use information about a specific alternative and test whether the null can predict the performance of the alternative. This procedure do the hypothesis testing that prior assumes one model is the true model against the competing model.

The other approach is a LR based test developed by Vuong (1989) which adopt the classical hypothesis testing framework and consider the Kullback-Leibler (1951) Information Criterion (KLIC) which measures the distance between a given distribution and the true distribution. Compares to Cox's approach, Vuong's approach does not assume a true model in advance. Instead, the two competing models are compared on the same basis.

In this chapter I explore the use of Vuong's model selection tests in the context of two competing hurdle, or two-tier, models. In particular, I am interested in models to explain Tobit-like outcomes: the response variable can be zero with positive probability, but it is continuous for strictly positive values. Such variables arise often out of economic optimization. Good examples include annual labor supply and amount of charitable contributions. For some nontrivial segment of the population, the optimum choice is zero; a so-called corner solution.

Two-tier models offer more flexibility than the most common model for corner solution applications: the standard, or Type I, Tobit model. Rather than the standard Tobit, they can specify different set of parameters for each tier, which is more economically reasonable. As an example of study that test between two-tier model and the standard Tobit model, Fin and Schmidt (1984) studied the two-tier model and developed a

test for the model selection between a two-tier model and standard Tobit model for nested models.

This chapter provides some tests and simulations that selecting alternative models for Tobit specification of censored regression models. Censored regression models are used when the dependent variable is partly continuous but piles up at one or more points with positive probability. We can group censored regression models into two categories: one is data censoring, the other is corner solution application. In the first category, there is a well-defined variable y^* and we are interested in the population regression $E(y^* | x)$ where x represents exogenous variables. If y^* and x were observed for everyone in the population then we can just apply ordinary least squares or nonlinear least squares. If y^* is censored above or below some value then a data problem arises. An example is when y^* is wage rate and wage is top-coded in the sampling survey: the actual value of wage rate is recorded up to some threshold like \$100, but after that only the fact wage rate was higher than \$100 is recorded. The second category of censored regression models is corner solution application, which appears more often in econometrics. Suppose y is an observable choice variable of an individual. If y is continuous then we can usually apply some standard procedures like ordinary least squares. But if y is something like individual consumption of a particular good, it will be optimal not to consume the particular good at all for some individuals so $y = 0$. Another example like workers decide whether goes to work or not and it is optimal for some workers not to work at all. In this kinds of case, the distribution of y in the population will be continuous for $y > 0$ but y is zero with positive

probability. The data observability is not the issue in this category of censored regression models. The focus is the expectation of y like $E(y | x)$ and $E(y | x, y > 0)$.

In this chapter we focus on the corner solution application category of censored regression models. As we said, the main feature of this kind of models is that the range of dependent variable basically is continuous yet has a lower bound, say $y = 0$, and observations pile up at this point. The fact that the observed outcome of y might be zero simply reflects the choice of an individual. If we assume $E(y | x) = x\beta$ and apply OLS on a random sample, then OLS estimate of β will be consistent if $E(y | x)$ is in fact linear. However, the predicted value for y can be negative for some combinations of x and β . Also, a transformation like $\log(y)$ does not work because $\log(0)$ is not defined. Actually, $E(y | x)$ is not linear in many cases. Since the usual regression model can not be used in this kind of situation, special treatment has to be applied.

It is well known that Tobit model was created to deal with the censored regression models. For the population, we can write the standard censored Tobit model as

$$y^* = x\beta + u, \quad u | x \sim \text{Normal}(0, \sigma^2) \quad (3.1)$$

$$y = \max(0, y^*) \quad (3.2)$$

For the case of true data censoring, $E(y^* | x)$ is of interest and the estimation of β is what we need. In the case of corner solution application, we are interested in $E(y | x)$ or $E(y | x, y > 0)$ which will depend on β nonlinearly instead of linearly depend on it.

However, a major limitation of Tobit model to handle corner solution application is that standard Tobit model uses the same mechanism to determine the choice between, say,

buying or not or working or not and the amount to buy or the amount to work. If we define y as the variable to be explained, that means Tobit model use only one set of procedure determining the choice between $y=0$ versus $y>0$ and the amount of y once $y>0$. That is not very reasonable because Tobit does not allow the determination of the amount of y when it is not zero to depend on different parameters or variables from those determining the probability of its being zero.

Some alternatives to the censored Tobit model have been suggested to improve the insufficiency of the standard censored Tobit model. These models allow the decision of $y>0$ versus $y=0$ to be separate from the decision of the amount of y given that $y>0$. These kinds of models are so called “two-tiered models” or “hurdle models”.

The model suggested by Cragg (1971) is one of the most widely used hurdle models. Cragg suggested a model that nests the usual Tobit model to form a truncated normal distribution. A major advantage of this model is that it can be easily transferred to standard Tobit model. Because of Tobit model is a special case of the model suggested by Cragg, it is possible to test Tobit model versus Cragg's alternative. Fin and Schmidt (1984) derive the LM test for this purpose.

There are other alternatives to the standard Tobit model that have been developed. A simpler lognormal distribution model is the most suggested one. Unlike Cragg's alternative that truncate normal distribution at zero, lognormal distribution model assumes that the dependent variable follows a lognormal distribution when the dependent variable is greater than zero. The model is easier to consistently estimate the expectation y conditional on explanatory variables than Cragg's alternative. However, this model does

not contain the standard Tobit model as a special case by imposing parameter restrictions, so it is difficult to test the Tobit model against this alternative.

There are some methods that can be applied to test which model should be selected. White and Olson (1979) evaluated competing models by their mean square error of prediction. Vuong (1989) basically compared the maximum likelihood values of the competing models' maximum likelihood functions to select one model over the other.

This chapter uses the maximum likelihood value comparison approach to test model selection between Cragg model and the lognormal one. We use three different data sets along with computer simulation to decide which one is better to fit data.

Section II discusses model specifications and model selection procedures. Section III contains the empirical work and results. Section IV is the conclusion.

II. Basic Framework

As we discussed in section I, the situation being considered can be divided into two tiers. First, we consider a particular event at each observation may or may not occur. If it does not occur, the variable has a zero value. If it does occur, thing goes into the second tier and the event associate with a continuous, positive random variable. We now define y is the variable to be explained. First we will examine Cragg model.

Cragg suggests the truncated normal distribution which truncating the distribution at zero in the second tier to ensure the dependent variable y is positive. We

assume
$$P(y = 0 | x) = 1 - \Phi(x\gamma), \quad (3.3)$$

where $\Phi(\cdot)$ is the standard normal cdf and γ is a $K \times 1$ vector of parameters. The density conditional on $y > 0$ is assumed to be

$$f(y | x, y > 0) = \{\Phi(x\beta/\sigma)\}^{-1} [\phi(\{y - x\beta\}/\sigma)/\sigma], \quad y > 0, \quad (3.4)$$

where x is a $1 \times K$ vector of the values of the independent variables, β is a $K \times 1$ vector of parameters, and $\phi(\cdot)$ is the standard normal pdf. The first equation shows the probability that y is zero or positive. The second equation is the density function of y when y is greater than zero. The density of y given x is

$$f(y | x; \theta) = \{1 - \Phi(x\gamma)\}^{1[y=0]} (\Phi(x\gamma) \{\Phi(x\beta/\sigma)\}^{-1} [\phi(\{y - x\beta\}/\sigma)/\sigma])^{1[y>0]};$$

Cragg's model nests the usual Tobit model by imposing the restriction

$$\gamma = \beta / \sigma.$$

The log-likelihood function of truncated normal model for observation i is:

$$\begin{aligned} \lambda_i(\theta) &= \log(f_i(y | x; \theta)) \\ &= 1[y_i = 0] \log\{1 - \Phi(x_i\gamma)\} \\ &\quad + 1[y_i > 0] \left\{ \log\Phi(x_i\gamma) - \log\Phi(x_i\beta/\sigma) + \log\phi\left(\frac{y_i - x_i\beta}{\sigma}\right) - \log\frac{\sigma^2}{2} \right\} \\ &= 1[y_i = 0] \log\{1 - \Phi(x_i\gamma)\} \\ &\quad + 1[y_i > 0] \left\{ \log\Phi(x_i\gamma) - \log\Phi(x_i\beta/\sigma) - \frac{1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (y_i - x_i\beta)^2 \right\} \quad \dots(3.5) \end{aligned}$$

As a competitor to Cragg's model, we assume that, conditional on $y > 0$, $\log(y)$ follows a normal distribution. Thus,

$$P(y = 0 | x) = 1 - \Phi(x\gamma) \quad (3.6)$$

$$\log(y) | x, y > 0 \sim \text{Normal}(x\beta, \sigma^2). \quad (3.7)$$

The first equation indicates the probability of y being zero or positive and is identical to Cragg's model. The second equation says that conditional on $y > 0$, $y|x$ follows a lognormal distribution. The density of y given x can be written as

$$f(y | x; \theta) = \{1 - \Phi(x\gamma)\}^{1[y=0]} \{\Phi(x\gamma)\phi[\{\log(y) - x\beta\}/\sigma]/(y\sigma)\}^{1[y>0]} \quad (3.8)$$

The log-likelihood function for observation i is

$$\begin{aligned} \lambda_i(\theta) &= \log(f_i(y | x; \theta)) \\ &= 1[y_i = 0] \log\{1 - \Phi(x_i\gamma)\} + 1[y_i > 0] \left\{ \log \Phi(x_i\gamma) - \log(y_i) - \frac{1}{2} \log(\sigma^2) \right. \\ &\quad \left. - \frac{1}{2} \log(2\pi) - \frac{1}{2} (\log(y_i) - x_i\beta)^2 / \sigma^2 \right\} \end{aligned} \quad (3.9)$$

Unlike Cragg's model, this does not nest the standard Tobit model. But it is easy to interpret. For example, β_j is the semi-elasticity of y with respect to x_j , conditional on $y > 0$. This makes the model easier to have the economic interpretation than Cragg's model. The expectations of $E(y | x)$ and $E(y | x, y > 0)$ are also different in two models. In lognormal model,

$$E(y | x, y > 0) = \exp(x\beta + \sigma^2 / 2), \quad E(y | x) = \Phi(x\gamma) \exp(x\beta + \sigma^2 / 2), \quad \text{and}$$

these are easily estimated given $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{\gamma}$.

Cragg's model and the lognormal model are not nested, so we cannot use standard tests. As discussed in the introduction, we could apply Cox's procedure to testing one model- which is assumed to be true under H_0 - against the other. For the two-tier models studied here, this is very complicated and computationally expensive. Young (1989) suggested a LR (likelihood ratio) based test which consider the Kullback-Leibler (1951) Information Criterion (KLIC) that measures the distance between a given

distribution and the true distribution. If the distance between a specified model and the true distribution is defined as the minimum of the KLIC over the distribution in the model, then the “best” model among a collection of competing models is the model that is closest to the true distribution (see, e.g., Young (1989)).

Suppose there are two competing models F_θ and G_γ that

$F_\theta = \{f(y | x; \theta); \theta \in \Theta\}$ and $G_\gamma = \{g(y | x; \gamma); \gamma \in \Gamma\}$, where F_θ and G_γ are conditional models and f and g are the conditional density functions. The models can be nested, non-nested, or overlapping, and the model selection test is to test the null hypothesis that $E^0[\log f(y | x; \theta_*)] = E^0[\log g(y | x; \gamma_*)]$ meaning that the two models are equivalent against $E^0[\log f(y | x; \theta_*)] > E^0[\log g(y | x; \gamma_*)]$ meaning that F_θ is better than G_γ or against $E^0[\log f(y | x; \theta_*)] < E^0[\log g(y | x; \gamma_*)]$ meaning that G_γ is better than F_θ .

Although the quantity $E^0[\log f(y | x; \theta_*)]$ and $E^0[\log g(y | x; \gamma_*)]$ are unknown, they can be consistently estimated by $(1/n)$ times the log-likelihood evaluated at the pseudo or quasi maximum likelihood estimator (MLE) (see, e.g., White (1982a)). It is apparent that $(1/n)$ times the log likelihood ratio (LR) statistic is a consistent estimator of $E^0[\log f(y | x; \theta_*)] - E^0[\log g(y | x; \gamma_*)]$. Cox (1961, 1962) and White (1982b) showed that, if n is the sample size, then $n^{-1/2}$ times the LR statistic properly centered and normalized has a limiting standard normal distribution under the hypothesis that one of the competing models is correctly specified.

To pick the better model and to derive the tests for model selection, the model with the minimum KLIC which measures the distance between the true distribution and a specified model should be choose. For the conditional model F_θ , KLIC is defined as

$$\text{KLIC}(H_{Y|X}^0; F_\theta) \equiv E^0[\log h^0(Y_t | X_t)] - E^0[\log f(Y_t | X_t; \theta_*)],$$

where $H_{Y|X}^0$ is the true conditional distribution of Y_t given X_t . $h^0(\cdot | \cdot)$ is the true conditional density of Y_t given X_t . Since the expression in the equation does not depend on F_θ , an equal measure is $E^0[\log f(Y_t | X_t; \theta_*)]$.

If we have two competing models, the model that is closest to the true conditional distribution should be chose. Vuong therefore develop the following hypotheses:

$$H_0 : E^0 \left[\log \frac{f(Y_t | X_t; \theta_*)}{g(Y_t | X_t; \gamma_*)} \right] = 0, \quad (3.10)$$

meaning that F_θ and G_γ are equivalent, against

$$H_f : E^0 \left[\log \frac{f(Y_t | X_t; \theta_*)}{g(Y_t | X_t; \gamma_*)} \right] > 0, \quad (3.11)$$

meaning that F_θ is better than G_γ , or

$$H_g : E^0 \left[\log \frac{f(Y_t | X_t; \theta_*)}{g(Y_t | X_t; \gamma_*)} \right] < 0, \quad (3.12)$$

meaning that F_θ is worse than G_γ .

Since the two competing models we are going to examine (the lognormal model and the Cragg alternative) are non-nested, the model selection tests for strictly non-nested models are:

$$H_0 : n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \xrightarrow{D} N(0,1) \quad (3.13)$$

$$H_f : n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \xrightarrow{a.s.} +\infty \quad (3.14)$$

$$H_g : n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \xrightarrow{a.s.} -\infty \quad (3.15)$$

where $LR_n(\hat{\theta}_n, \hat{\gamma}_n) \equiv L_n^f(\hat{\theta}_n) - L_n^g(\hat{\gamma}_n) = \sum_{i=1}^n \log \frac{f(y_i | x_i; \hat{\theta}_n)}{g(y_i | x_i; \hat{\gamma}_n)}$, where $\hat{\theta}_n$ and $\hat{\gamma}_n$ are the maximum likelihood estimators of θ_* and γ_* . Also, $L_n^f(\hat{\theta}_n) = \sup_{\theta \in \Theta} L_n^f(\theta)$ and $L_n^f(\theta)$ is the conditional log-likelihood function for the model F_θ . A similar definition applies to $\hat{\gamma}_n$ for the conditional model G_γ . And $\hat{\omega}_n = \left\{ \frac{1}{n} \sum_{i=1}^n \left[\log \frac{f(y_i | x_i; \hat{\theta}_n)}{g(y_i | x_i; \hat{\gamma}_n)} \right]^2 \right\}^{1/2}$.

The tests above provide us a useful framework for model selection. We can choose a critical value c from the standard normal distribution at some significance level. If the value $n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ is greater than c , then the model F_θ is preferred. If the value $n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ is smaller than $-c$, then the model G_γ is preferred. If the value $n^{-1/2} LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ is between c and $-c$, then we cannot discriminate between the two competing models.

III. Empirical Result

To determine the model selection between lognormal model and Gragg's alternative, we divide this section into two parts. In the first part we do a simulation using computer-generated data sets to see how well the test works. In the second part we apply

the test introduced in section II to three data sets used in chapter one and two. We basically follow the null hypothesis:

$$H_0 : \frac{\sum_{i=1}^n (\log \hat{f}_i - \log \hat{g}_i)}{\sqrt{\sum_{i=1}^n (\log \hat{f}_i - \log \hat{g}_i)^2}} \stackrel{a}{\sim} N(0,1)$$

to form the test and apply to different data sets.

(I) The Simulation Results

The simulation is designed as the following:

We use computer to generate data sets. The data sets are generated in two kinds: one follows lognormal distribution and the other follows truncate normal distribution.

(a) lognormal distribution

The data sets follow lognormal distribution is generated by:

(i) Generate $X = (1, X_1, X_2)$,

where 1 is constant term, X_1 is a binary variable which is zero for around 70% and one for around 30%. X_2 is a standard normal random variable.

(ii) Generate $z = 1$ if $X\gamma + u \geq 0$, otherwise $z = 0$. u follows a standard normal distribution.

(iii) If $z = 0$ then $y = 0$. If $z = 1$, then $y = \exp(X\beta + v)$.

v is generated to corrected with u such that

$v = .3u + .7k$ where k follows a normal distribution such that $k \sim N(0, \sigma^2)$.

(b) truncate normal distribution

(i) Follow the steps (i) and (ii) in (a).

(ii) If $z = 0$ then $y = 0$. If $z=1$, then $y = X\beta + v$ is truncated at zero. Also, v is generated to correlated with u such that $v = .3u + .7k$ where k follows a normal distribution such that $k \sim N(0, \sigma^2)$.

With both kinds of the data sets, we set sample size at 200 and 1000. We also experiment two sets of true values of r and β^5 along with two σ^2 values (.2 and .5). To see if the simulation results change if we increase the variation of X_2 , we also experiment the whole simulation by doubling X_2 's variation. The whole combinations show in Table 3.1. There are sixteen combinations for each kind of data set. With all the combinations, we run each of them 200 times to find out the pattern.

We choose 2.58 as the critical value so the significant level is 1%. If the statistic value we calculate is greater than 2.58, then lognormal model is preferred. If the statistic value is less than -2.58, then Gragg truncate normal model is preferred. If the statistic value is between 2.58 and -2.58, then there is no conclusion which model is better. The rejection rate for the true model being picked up correctly shows in Table 3.2 and 3.3.

In Table 3.2, the true model is lognormal model. We have the following findings:

(1) The rejection rates decrease when we decrease the sample size from 1000 to 200. This is reasonable since bigger sample should give the test more power to distinguish the true model from the alternative model.

⁵ The first set is: $\gamma = (0,1,1)$ and $\beta = (0,1,1)$. The second set is: $\gamma = (0,-1,1)$ and $\beta = (0,1,1)$.

- (2) We also find the rejection rates decrease when σ^2 increases from .2 to .5. That means the test is harder to pick up the right model when the variance of the error term in equation y increases.
- (3) The rejection rates decrease when the variation of explanatory variable increases. That means the test is less powerful to pick up the right model when the variation of the explanatory variable increases.

In Table 3.3, the true model is Cragg's truncate normal model. The same test is used as we did in Table 3.2. We have the following findings:

- (1) The rejection rates decrease as we decrease the sample size from 1000 to 200. This result is similar with Table 3.2's result.
- (2) The rejection rates do not change with respect to σ^2 and the variation of X_2 change. It's more stable in our simulation.
- (3) The specification that $\gamma = (0, -1, 1)$ and $\beta = (0, 1, 1)$ may not be a good one when the true model is Cragg's truncate normal model. The maximum likelihood procedure just can not converge when estimating the log likelihood function for Cragg's truncate normal model.

Compare the simulation results for two different kinds of models, we noticed that Cragg's truncate normal model is more robust than the lognormal model in the sense that the model is seldom being wrong picked when we change the variation of X and σ^2 . Only when we decrease the sample size the model's rejection rates decrease. This means the possibility for us to pick up a wrong model is higher when the true model is a lognormal model.

(II) The Real Data Sets Results

We use the 3 data sets at hand to run the test we described above to see which model is better fitted. The 3 data sets we use are exactly the data sets we used in previous two chapters. The first one is Moffitt data set. The second one is PSID data set and the third one is CPS data set.

In the case of Moffitt's data set, we use hours of work a week as the dependent variable. The independent variables include dummies for marital status, age, race, the number of family members, the number of children in the household who are less than 6, the number of children in the household who are greater than 6, years of education, the size of the local labor force, the employment fractions in manufacturing and in government in the census region of residence and the flow of the asset income. After conducting the maximum likelihood procedures for both models, the calculated statistic value is -8.10 . Cragg's truncated normal model is preferred.

For PSID data set, the hour of work is the dependent variable. The explanatory variables include age, non-wife income, and the number of kids less than age of six, the number of kids greater than age of six, years of education, experience, and square of experience. The statistic value is -6.8 . The Cragg's truncate normal model is preferred.

For CPS data set, the dependent variable is hours of work. The explanatory variables are age, non-wife income, race (a dummy variable), education, experience, square of experience, husband education, husband experience, the number of kids less than age of six, the number of kids greater than age of six. The statistic value is -11.87 . Again, The Cragg's truncate normal model is preferred.

IV. Conclusion

In this chapter, we discuss the potential drawbacks of the Tobit model to handle the corner solution applications of censored regression models. We also have two good alternative candidates for Tobit model. One is a lognormal model and the other one is Cragg's truncated normal model. To choose the model that best fits the data, we applied Young's LR based test for model selection.

We do simulation first and then apply the test to the data sets we used in chapter one and two. The simulation results show that it is easy for us not to pick the lognormal model even when the lognormal model is the true model. So if the test shows that lognormal model is preferred, the chance to go wrong is very small. On the other hand, if the test indicates Cragg's truncate normal model is preferred, we still have to be careful.

The results for the real data sets show that Cragg's truncate normal model is preferred in all three data sets although we should be more careful about the result as the simulation suggest.

Table 3.1

Combinations of the Simulation

Sample Size: 1000				Sample Size: 200			
$X_2 \sim N(0,1)$		$X_2 \sim 2^* N(0,1)$		$X_2 \sim N(0,1)$		$X_2 \sim 2^* N(0,1)$	
$\sigma^2 = .2$	$\sigma^2 = .5$	$\sigma^2 = .2$	$\sigma^2 = .5$	$\sigma^2 = .2$	$\sigma^2 = .5$	$\sigma^2 = .2$	$\sigma^2 = .5$
$\gamma = (0,1,1)$							
$\beta = (0,1,1)$
$\gamma = (0,-1,1)$							
$\beta = (0,1,1)$

Table 3.2

The Rejection Rate

True Model: Log Normal Model

Note: Critical value is 2.58

Sample Size: 1000					Sample Size: 200			
	X~ N(0,1)		X~2* N(0,1)		X~ N(0,1)		X~2* N(0,1)	
	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$
$\gamma=(0,1,1)$								
$\beta=(0,1,1)$	100%	98%	56.25%	32.46%	100%	92.5%	27.37%	12.35%
	(200)	(200)	(192)	(191)	(200)	(200)	(179)	(170)
$\gamma=(0,-1,1)$								
$\beta=(0,1,1)$	100%	99.5%	100%	100%	100%	99%	100%	100%
	(200)	(200)	(153)	(136)	(200)	(200)	(191)	(190)

Table 3.3

The Rejection Rate

True Model: Truncated Normal Model

Critical value is -2.58

Sample Size: 1000					Sample Size: 200			
	X~ N(0,1)		X~2* N(0,1)		X~ N(0,1)		X~2* N(0,1)	
	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$	$\sigma^2=.2$	$\sigma^2=.5$
$\gamma=(0,1,1)$								
$\beta=(0,1,1)$	100%	100%	100%	100%	93%	87.5%	100%	99%
	(200)	(200)	(200)	(200)	(200)	(200)	(200)	(200)
$\gamma=(0,-1,1)$								
$\beta=(0,1,1)$	*	*	*	*	*	*	*	*

note: * means the computer work does not converge

Chapter 4

CONCLUSION

In the previous three chapters, we explore several different topics for Tobit model that takes the form

$$y_1 = z_1\gamma + u_1 \quad (4.1)$$

$$y_2 = \max(0, \alpha_0 y_1 + z_2\delta + u_2) \quad (4.2)$$

where equation (4.1) is the structural equation and equation (4.2) is the Tobit selection equation. z_1, z_2 are the vectors of exogenous variables.

The first two chapters concentrate on equation (4.1)'s consistent estimation. In chapter one we derive and apply an alternative method to test and correct sample selection bias while estimating essentially the Type III Tobit model with endogenous explanatory variables in the structural equation. We applied multi-step approach to estimate traditional Type III Tobit model which usually being estimated by Maximum Likelihood Method. Multi-step approach has advantages of easy to compute and more robust than MLE does. It also can provide an easy t-test for sample selection bias at the same time. However, if one need to acquire the asymptotic variance matrices under this multiple step approach, it maybe needs a lot of computational efforts.

The NLS data set used by Moffitt (1984) is applied to our procedure. We find there are evidence of the existence of sample selection bias in our estimation, which is different from what Moffitt's claim. After correcting the bias, the average return of education goes from 7.77% to 5.94% which is similar to the early findings. We can say that the average return will be overestimate if we don't correct the sample selection bias.

The coefficient of hours^* also changes from .36% to 1.21% before and after we add the constructed variable \hat{v}_2 into the structural equation. It means that the hours effect to wage will be underestimated if we don't take the sample selection bias into account.

In chapter two we focus on the situation where there are both self-selection and sample selection. We show how to test and to correct for self-selection and sample selection biases at the same time, especially in circumstance when the explanatory variable, such as education, interacted with unobservable. Like in chapter 1, we derive a multiple-step procedure which provides an easier to compute, less assumptions needed alternative to traditional Maximum Likelihood Estimation. The procedure extends Garen's (1984) procedure to handle the sample selection problem besides the self-selection problem. Two data sets are applied to this procedure. One is PSID data set of 1975 used by Mroz (1987) in his study of the sensitivity of female labor supply to various assumptions. The other one is CPS data set of May 1991 compiled by Daniel Hamermesh.

The empirical result for PSID data set shows that there is evidence of self-selection biases but not strong evidence of sample selection biases. The simple OLS regression using only women who work cannot produce consistent estimates and the estimate of return to education is biased upward. By contrast, 2SLS can provide consistent estimates when sample selection bias is not a problem.

In CPS data set, both self-selection biases and sample selection biases exist. The method we suggested should be used to correct self-selection biases along with sample selection biases. The OLS estimate for returns to education is biased downward. This indicates there is no certain direction for OLS biases.

Chapter three shift the focus to equation (4.2). We discussed the possible model misspecification and explore the model selection test for choosing between two competing models. The potential drawbacks of the Tobit model to handle the corner solution applications of censored regression models are also being discussed. There are two alternatives for Tobit model. One is lognormal model, the other is Cragg's truncate normal model. To pick up a right one that fits the data better, Vuong's LR based test for model selection is applied.

The simulation results show that the Vuong's model selection test basically can pick up the correct model. However, we do find that using the test is easier to mispick when the lognormal model is the true model. So if the test shows that lognormal model is preferred, the chance to go wrong is very small. On the other hand, we still have to be careful if the test indicates Cragg's truncate normal model is preferred.

For further research, in chapter one and two the normal distribution assumption for error term in the selection equation could be relaxed. Powell (1984) extends least absolute deviations (LAD) estimation to the regression with non-negativity of the dependent variable, and gives conditions under which this estimator is consistent and asymptotically normal. It should be interesting to explore and apply LAD to our case. In chapter three, the model selection test is under the assumption that there are only two competing models. It is useful to generalize the procedures to the case where there are many competing models.

REFERENCES

- Amemiya, T. (1985), *Advance Econometrics*. Cambridge: Harvard University Press.
- Angrist, J. D., and A. B. Krueger (1991). "Does compulsory school attendance affect schooling and earnings?", *Quarterly Journal of Economics* v106, n4, p979(36).
- Annenfelter, O., and A. B. Krueger (1992) "Estimates of the Economic Return to Schooling for a New Sample of Twins." Princeton University Industrial Relations Section Working Paper #304.
- Butcher, K. F., and A. Case (1993) "The Effect of Sibling Composition on Women's Education and Earnings." Unpublished Discussion Paper, Princeton University Department of Economics.
- Card, D., (1993). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling" NBER Working Paper No. 4483.
- Garen, J., (1984). "The returns to schooling: a selectivity biases approach with a continuous choice variable, *Econometrica* 52, 1199-1218.
- Chow, G. (1983): *Econometrics*. New York: McGraw-Hill .
- Cox, D. R. (1961): "tests of Separate Families of Hypotheses," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 105-123.
- Cox, D. R. (1962): "Further Results on Tests of Separate families of Hypotheses," *Journal of Royal Statistical Society, Series B*, 24, 406-424.
- Cragg, J. (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," *Econometrica* 39,829-844.

- Fin, T. and P. Schmidt (1984), "A Test of the Tobit Specification Against an Alternative Suggested by Cragg," *Review of Economics and Statistics* 66, 174-177.
- Gronau, R.(1974). "Wage Comparisons—A Selectivity Biases." *Journal of Political Economy* 82:1119-43
- Hausman, J. (1980), "The Effects of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation." *Journal of Public Economics* 14:161-94.
- Heckman, J. (1974). "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42:679-94
- Heckman, 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* 5:475-92.
- Kane, T. J., and C. E. Rouse (1993) "Labor Market Returns to Two- and Four- Year Colleges: Is a Credit a Credit and Do Degrees Matter?" Princeton University Industrial Relations Section Working Paper #311.
- Kullback, S., and R. A. Leibler (1951): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79-86.
- Larson, D. (1979), "Taxes in a Labor Supply Model with Joint Wage-Hours Determination: Comment." *Econometrica* 47:1311-13.
- Lee, L. F., and R. P. Trost (1978). "Estimation of Some Limited Dependent Variable Models with Application to Housing Demand." *Journal of Econometrics* 8:357-82.
- Lee, L.F. (1978). "Unionism and Wage Rate: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19:415-33.

- Lewis, H. G. (1974). "Comments on Selectivity Biases in Wage Comparisons." *Journal of Political Economy* 82(6):1145-55.
- Moffitt, R. (1984), "The Estimation of a Joint Wage-Hours Labor Supply Model" *Journal of Labor Economics* 2:550-566.
- Mroz, T.A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica* 55, 765-799.
- Oi, W. (1962), "Labor as a Quasi-fixed Factor." *Journal of Political Economy* 70:538-55.
- Powell, J.L. (1984), "Least Absolute Deviations Estimation for the Censored regression Model" *Journal of Econometrics* 25:303-325.
- Rosen, H. (1976), "Taxes in a Labor Supply Model with Joint Wage-Hours Determination." *Econometrica* 44:485-507
- Rosen, S. (1969), "On the Interindustry Wage and Hours Structure." *Journal of Political Economy* 77:249-73.
- Roy, A. D. (1951). "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3:135-46.
- Vella, F. (1992), "Simple Tests for Sample Selection Bias in Censored and Discrete Choice Models," *Journal of Applied Econometrics* 7,413-421.
- Vella, F. (1993), "A Simple Estimator for Simultaneous Models with Censored Endogenous Variables," *International Economic Review* 34, 441-457.
- Vuong, Q. (1989), "Likelihood Ratio Tests for Model Selection and Nonnested Hypotheses," *Econometrica* 57, 307-333.
- White, H., and L. Olson (1979): "Determinants of Wage Change on the Job: A symmetric Test of Non-Nested Hypotheses," mimeo, University of Rochester.

- White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.
- White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-25
- White, H. (1982): "Regularity Conditions for Cox's Test of Non-Nested Hypotheses," *Journal of Econometrics*, 19, 301-318.
- White, H. (1982), "Instrumental Variables Regression with Independent Observations," *Econometrica* 50, 483-500.
- Willis, R., and S. Rosen (1979). "Education and Self-Selection." *Journal of Political Economy* 87(5, Part 2):507-36.
- Wooldridge, J.M. (1996), "Selection Corrections with a Censored Selection Variable" Working Paper, Michigan State University.
- Wooldridge, J.M. (1997), "On Two Stage Least Squares Estimation of the Average Treatment Effect in a Random Coefficient Model", Working Paper, Michigan State University.
- Wooldridge, J.M. (1998), "Econometric Analysis of Cross Section and Panel Data", Manuscript, Michigan State University.

MICHIGAN STATE UNIV. LIBRARIES



31293017014741