ESTIMATING NONLINEAR CROSS SECTION AND PANEL DATA MODELS WITH ENDOGENEITY AND HETEROGENEITY

by

Hoa Bao Nguyen

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

ECONOMICS

2011

ABSTRACT

ESTIMATING NONLINEAR CROSS SECTION AND PANEL DATA MODELS WITH ENDOGENEITY AND HETEROGENEITY

by

Hoa Bao Nguyen

The dissertation consists of three chapters that consider the estimation of nonlinear cross section and panel data models. This study contributes to the literature by developing new estimation methods for estimating models with limited dependent variable and endogenous regressors in the presence of unobserved heterogeneity. It also makes contribution to the field of labor economics by applying my new estimators to the study of female labor supply.

In the first chapter, a fractional response model with a count endogenous regressor is considered. A new estimation method is proposed to handle discrete endogeneity in the presence of unobserved heterogeneity and non-linear setting. The two-step Quasi-Maximum Likelihood and Nonlinear Least Squares estimators using the Adaptive Gauss Hermite quadrature are proposed. Average partial effects for discrete endogenous variables are obtained given its difficulty of approximation based on a non-closed form conditional mean with a non-normal heterogeneity. Monte Carlo simulations verify that the new estimators are the least biased and the most efficient among examined estimators including existing estimators. This is the first research that supports the necessity and significance of count endogeneity. The proposed estimators are applied to analyze the US female labor supply. The result shows diminishing marginal effects of additional children on female's working hours. This novel finding is consistent with a story of fertility and presents an evidence of economies of scale that mothers become more efficient after raising the first kids, devote more time to work and balance between working time and family time.

In the second chapter, a dynamic Tobit panel data model that allows for an endogenous regressor (besides the lagged dependent variable) is developed. I also permit the presence of unobserved heterogeneity and serial correlation of transitory shocks. A correlated random effect Tobit approach, a computationally attractive estimation method, is proposed. The estimation method employs the control function approach to account for endogeneity and to consistently estimate average partial effects. In addition, serial correlation in the reduced form is corrected which makes the estimator more robust. This method is readily applied to Panel Study of Income Dynamics data from 1980 to 1992. I find a strong evidence of persistence in US white female labor working hours and the initial condition of female labor supply is statistically significant.

The third chapter considers the estimation of a panel data model with a corner solution response and the presence of a dummy endogenous variable as well as heterogeneity. The main contribution is to allow a joint distribution of the binary endogenous regressor and the unobserved factors that affect both the amount and participation equations. A bivariate probit model is suggested in the first stage. An exponential type II Tobit (ET2T) model is exploited for the amount equation to ensure that the predicted value for the response variable is positive; and there is a correlation between unobserved effects in both the amount and participation equations. The two-step estimation procedure inspired by Heckman's idea of adding correction terms for endogenous switching and a corner solution outcome is used to analyze the impact of fertility on female labor force participation and labor supply using the Vietnamese Household Living Standard Surveys data 2004-2008. The proposed approach gives a statistically significant negative effect of having a newborn on women who are working and remain in the labor market. It corrects remarkably the bias in estimating the effect of a newborn on mother's working hours compared to other alternative estimation methods.

Copyright by Hoa Bao Nguyen 2011 This thesis is dedicated to my family, my husband, Minh Cong Nguyen, and my son, Ton Chi Cong Nguyen.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank people who have helped me during the journey to a Ph.D. First, I would like to express my deepest gratitude to my advisor Professor Jeffrey Wooldridge, a person of great knowledge and exceptional teacher, for his generous advice, support and excellent training during my work on this dissertation.

I would also like to thank my other committee members, Professors Peter Schmidt, Todd Elder and Joseph Gardiner for their valuable comments and support.

I am very grateful for the support that I receive from people at the World Bank who kindly encouraged me to apply and develop more econometric models and estimation methods for nonlinear panel data with discrete endogenous variables, to be exploited as robust devices in useful applications.

I wish to thank faculty members and graduate students of the Department of Economics at Michigan State University for their useful training of many core economics branches and seminar discussion of econometrics topics.

I would like to express a warmest gratitude to my parents and my husband. I owe my father because he has motivated me to be a scientist and always encouraged as well as challenged me to make great accomplishments. I am grateful to my mother and my parents-in-law who have been there for me and make the time for me to focus on my dissertation. I was fortunate to have a beloved husband who has helped me continuously and tremendously during my doctorate. Without his love and support, I will never make it through. I also want to thank my sister and other members in my extended family. Last but not least, I wish to thank my baby, Tony, since having him during the graduate study made me recognize many values of life and created my unstoppable determination to obtain the PhD and other future accomplishments. I dedicate this dissertation to my big family.

I would like to thank everyone whom I did not mention specifically but who helped me during my studies.

TABLE OF CONTENTS

LIST O	F TAB	LES	ix
LIST O	F FIGU	JRES	xi
CHAPTER 1		ESTIMATING A FRACTIONAL RESPONSE MODEL WITH A COUNT	
1 1	т. 1	ENDOGENOUS REGRESSOR	1
1.1	Introd		1
1.2	1 neore	Entimation Describer	4
	1.2.1	Estimation Procedure	9 10
	1.2.2	1.2.2.1 The Case with Exogenous Covariates and a Normally Distributed	10
		Heterogeneity	10
		1.2.2.2 The Case with a Count Endogenous Covariate and a Non-normally	
		Distributed Heterogeneity	11
1.3	Monte	Carlo Simulations	13
	1.3.1	Estimators	14
	1.3.2	Data Generating Process	14
	1.3.3	Experiment Results	16
		1.3.3.1 Simulation Result with a Strong Instrumental Variable	16
		1.3.3.2 Simulation Result with a Weak Instrumental Variable	19
		1.3.3.3 Simulation Result with Different Sample Sizes	20
	124	1.3.3.4 Simulation Result with a Misspecified Distribution	21
1 /	1.3.4	Conclusion from the Monte Carlo Simulations	21
1.4	Applic		22
1.5	Concil		29
СНАРТ	TER 2	ESTIMATION OF A DYNAMIC TOBIT PANEL DATA WITH AN	
		ENDOGENOUS VARIABLE AND AN APPLICATION TO FEMALE	
		LABOR SUPPLY	30
2.1	Introd	uction	30
2.2	Literat	ure Review	31
2.3	Model		35
2.4	Averag	ge Partial Effects	40
2.5	Serial	Correlation Correction	41
	2.5.1	Estimation Procedure	44
	2.5.2	Average Partial Effects	45
	2.5.3	Comparison	45
2.6	Empir	ical Example	46
	2.6.1	Data	47
	2.6.2	Estimation and Result	49
2.7	Conclu	usion	51

	TER 3 AN EXPONENTIAL TYPE II TOBIT PANEL DATA MODEL WITH	
	BINARY ENDOGENOUS REGRESSOR - APPLICATION TO ESTI-	
	MATING THE EFFECT OF FERTILITY ON MOTHERS' LABOR	
	FORCE PARTICIPATION AND LABOR SUPPLY	53
3.1	Introduction	53
3.2	Literature Review	55
3.3	Model and Estimation	58
3.4	Average Partial Effect	63
3.5	Empirical Example	65
	3.5.1 Overview of Data	65
	3.5.2 Estimation and Result	67
3.6	Conclusion	71
APPEN	DIX A TABLES FOR CHAPTER 1	73
APPEN	DIX B TABLES AND FIGURES FOR CHAPTER 2	90
APPEN	DIX C TABLES FOR CHAPTER 3	99
APPEN	DIX D TECHNICALITIES FOR CHAPTER 1	103
D 1		100
D.1	Details of the QML Estimator	103
D.1	Details of the QML EstimatorDetails of the QML EstimatorD.1.1Asymptotic Variance for the Two-step Estimator	103 103
D.1	Details of the QML Estimator	103 103 107
D.1 D.2	Details of the QML EstimatorDetails of the QML EstimatorD.1.1Asymptotic Variance for the Two-step EstimatorD.1.2Asymptotic Variance for the APEsDetails of the Tobit Model's Estimators	103 103 107 110
D.1 D.2 D.3	Details of the QML Estimator	103 103 107 110 113
D.1 D.2 D.3 D.4	Details of the QML EstimatorD.1.1Asymptotic Variance for the Two-step EstimatorD.1.2Asymptotic Variance for the APEsDetails of the Tobit Model's EstimatorsFormula of the NLS estimationDerivation of the Heterogeneity Distribution	103 103 107 110 113 114
D.1 D.2 D.3 D.4 APPEN	Details of the QML Estimator	103 103 107 110 113 114 115
D.1 D.2 D.3 D.4 APPEN E.1	Details of the QML Estimator	103 103 107 110 113 114 115 115
D.1 D.2 D.3 D.4 APPEN E.1 E.2	Details of the QML Estimator	103 103 107 110 113 114 115 115 118
D.1 D.2 D.3 D.4 APPEN E.1 E.2 APPEN	Details of the QML Estimator	103 103 107 110 113 114 115 115 118 120
D.1 D.2 D.3 D.4 APPEN E.1 E.2 APPEN F.1	Details of the QML Estimator	 103 103 103 107 110 113 114 115 115 118 120 120
D.1 D.2 D.3 D.4 APPEN E.1 E.2 APPEN F.1 F.2	Details of the QML Estimator	 103 103 107 110 113 114 115 118 120 120 122

LIST OF TABLES

A.1	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)	74
A.2	Simulation Result of the Coefficient Estimates (N=1000, $\eta_1 = 0.5$, 500 replications).	75
A.3	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.1$, 500 replications)	76
A.4	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.9$, 500 replications)	77
A.5	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications, $\delta_{23} = 0.3$)	78
A.6	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications, $\delta_{23} = 0$)	79
A.7	Simulation Result of the Average Partial Effects Estimates (N=100, $\eta_1 = 0.5$, 500 replications)	80
A.8	Simulation Result of the Average Partial Effects Estimates (N=500, $\eta_1 = 0.5$, 500 replications)	81
A.9	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)	82
A.10	Simulation Result of the Average Partial Effects Estimates (N=2000, $\eta_1 = 0.5$, 500 replications)	83
A.11	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, a1 is normally distributed, 500 replications)	84
A.12	Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)	84
A.13	Comparison of analytical and bootstrapping mean of standard errors (N=1000, $\eta_1 = 0.5, 200$ replications)	85
A.14	Frequencies of the Number of Children	86
A.15	Descriptive Statistics	86
A.16	First-stage Estimates using Instrumental Variables	87

A.17	'Estimates Assuming Number of Kids is Conditionally Exogenous	88
A.18	Estimates Assuming Number of Kids is Endogenous	89
B .1	Summary Statistics	90
B.2	Determinants of Female Working Experience - First stage regressions	91
B.3	Estimating Dynamic Female Labor Supply, Second Stage Regressions, Experience is Treated as an Endogenous Variable	92
B .4	Average Partial Effects on Female Labor Supply	93
C.1	Summary Statistics for the Whole Sample	99
C.2	Summary Statistics for Each Year in the Panel	100
C.3	Bivariate Probit Estimates of Fertility and LFP in the First Stage	101
C.4	Estimates for Log(Female Working Hours) Equation	102

LIST OF FIGURES

B .1	Distribution of Women's Annual Hours of Work in 1980-1992	94
B.2	Hours of Work vs. Experience	95
B.3	Hours of Work vs. Number of Children 0-2	96
B.4	Hours of Work vs. Number of Children 3-5	97
B.5	Hours of Work vs. Number of Children 6-17	98

Chapter 1

ESTIMATING A FRACTIONAL RESPONSE MODEL WITH A COUNT ENDOGENOUS REGRESSOR

1.1 Introduction

Many economic models employ a fraction or a percentage, instead of level values, as a dependent variable. In these models, economic variables of interest occur in fractions such as employee participation rates in 401(k) pension plans, firm market shares and fractions of total weekly hours spent working. These fractional response variables take values in the unit interval [0,1], which have both continuous and discrete characteristics. As suggested in Papke & Wooldridge (1996, 2008), we can model fractional response variables based on a correctly specified conditional mean and use a simple quasi-maximum likelihood estimator or nonlinear least squares (QMLE/NLS) method with the Bernoulli distribution. This method is more attractive than other standard approaches such as the MLE method with beta distribution or the log-odd transformation because it will give a direct estimate of the original dependent variable and ensure that the predicted value is in the unit interval.

Fractional response models (FRMs) with continuous or binary endogenous variables have been studied (see more in Papke & Wooldridge (2008) and Wooldridge (2010)). However, there has not been any well-developed estimation method and procedure to deal with count endogeneity in FRMs. Traditionally, a count endogenous explanatory variable (CEEV) is treated as a continuous endogenous variable and it is written in a linear fashion of covariates including instruments and additive error. A common approach such as the two stage least squares (2SLS) using the linear approximation always gives a constant marginal effect. This approach ignores the fact that the marginal effect of having one more unit of the CEEV on the outcome of interest might be more or less than the marginal effect of having the previous unit on the outcome. In order to acknowledge this fact, we should study FRMs and count endogeneity with the nonlinear approximation in the first stage. Specifically, we can handle a CEEV by allowing a Poisson distribution of the count variable in the reduced form. The heterogeneity term in the Poisson model is assumed to be correlated with the error term in the structural conditional mean. It is standard to allow the heterogeneity to follow a gamma distribution (which leads to the gamma error in the reduced form) because it results in a closed form solution, the Negative Binomial (NB) model. The key to correct for the endogeneity problem in this case is how we are willing to make an assumption on the joint distribution of errors. One strategy is to allow a linear correlation between the transformation of the gamma, which is now normally distributed, and the error in the structural conditional mean (see further discussion in Weiss (1999)). However, this assumption does not allow a direct relationship between the two errors which governs the endogeneity problem. The choice of the transformation function is the inverse of the standard normal and depends on unknown parameters in the distribution of the heterogeneity of the Poisson model. This strategy can be used to test for endogeneity of the count explanatory variable but it will make the evaluation of the likelihood function more computational and obtaining the conditional maximum likelihood estimator as well as its asymptotic covariance matrix is nontrivial and time-consuming. An alternative strategy is that we still allow the gamma heterogeneity in the Poisson model and a linear, direct correlation of this heterogeneity and the error term in the structural conditional mean. We then need to integrate out this heterogeneity in the structural conditional mean. As discussed in Winkelmann (2000), the heterogeneity in a Poisson model can be presented in terms of an additive correlated error or a multiplicative correlated error. However, the multiplicative correlated error has some advantage over the additive correlated error on grounds of consistency. As a result, a multiplicative correlated error is used in this model.

Because nonlinearity is allowed in both the reduced form and the main structural equations, a two-step estimation procedure is more attractive. In a simultaneous nonlinear equations model with a count dependent variable and a binary endogenous variable, Terza (1998) proposed a twostage method estimation method using a joint normal distribution of the error terms. He did not carry out the estimation of the Poisson Full-Information Maximum Likelihood (FIML) or explore its properties even though this approach was introduced in his paper. This is due to the computation burdensome of the FIML estimator. It emphasizes the advantage of the two-step estimation procedure that we can employ in this model. However, the joint normal distribution of the error terms is no longer appropriate in this case because a normal error term in the Poisson model will not lead to a closed form solution. Therefore, we have discussed the strategy on assuming the error terms as above. Besides an easier computational task, the two-step estimation procedure ensures that the predicted value lying in the rational range. Moreover, we do not need to find a conditional probability for each value of the CEEV without knowing in advance specific values of a general count explanatory variable, and avoid computing a conditional MLE which must be very difficult.

Other estimation method for a FRM with a CEEV can be considered. For example, semiparametric and nonparametric method can be used (see more in Das (2005)). However, this approach does not give estimates of the partial effects or the average partial effects of interest. If we are interested in estimating both parameters and average partial effects (APEs), a parametric approach will be preferred. In addition, in a nonlinear model, the quantity of interest is the APE which can be comparable to a linear model's estimate. Therefore, it is necessary and useful for applied economists and practitioners to obtain the APE and use the parametric model.

In this chapter, I show how to specify and estimate FRMs with a CEEV and an unobserved heterogeneity. Based on the work of Papke & Wooldridge (1996, 2008), I also use models for the conditional mean of the fractional response in which the fitted value is always in the unit interval. I focus on the probit response function since the probit mean function is less computationally demanding in obtaining the average partial effects. I suggest a new estimation method to handle discrete endogeneity in the presence of unobserved heterogeneity and non-linear setting. The two-step Quasi-Maximum Likelihood and Nonlinear Least Squares estimators using Adaptive Gauss Hermite quadrature are proposed. Average partial effects for discrete endogenous variables are obtained given its difficulty of approximation based on a non-closed form conditional mean with a non-normal heterogeneity. Monte Carlo simulations verify that the new estimators are the least biased and the most efficient among examined estimators including existing estimators. Using

these robust and efficient estimators, I applied my proposed estimators to analyze the US female labor supply. The empirical result gives an evidence to show that this is the first research that supports the necessity and significance of count endogeneity.

This chapter is organized as follows. Section 2 introduces the specifications and estimations of a FRM with a CEEV and shows how to estimate parameters and the average partial effects using the two-step QMLE and NLS approaches. Section 3 presents Monte Carlo simulations and an application to the fraction of total working hours for a female per week will follow in Section 4. Section 5 concludes.

1.2 Theoretical Model - Specificatio and Estimation

For a $1 \times K$ vector of explanatory variables z_1 , the conditional mean model is expressed as follows:

$$E(y_1|y_2, \mathbf{z}, a_1) = \Phi(\alpha_1 y_2 + \mathbf{z}_1 \delta_1 + \eta_1 a_1),$$
(1.1)

where $\Phi(\cdot)$ is a standard normal cumulative distribution function (cdf), y_1 is a response variable $(0 \le y_1 \le 1)$, and a_1 is a heterogeneous component or an omitted factor assumed to be correlated with y_2 but independent of exogenous variables z. In equation (1.1), I focus on the fractional probit conditional mean because it gives a computationally simple estimator when we deal with unobserved heterogeneity and endogenous regressors, as well as a convenient way to obtain average partial effects later on. The exogenous variables are $z = (z_1, z_2)$ where we need exogenous variables z_2 to be excluded from (1.1). z is a $1 \times L$ vector where L > K, z_2 is a vector of instruments. y_2 is a count endogenous variable where we assume that the endogenous regressor has a Poisson distribution:

$$y_2|\mathbf{z}, a_1 \sim Poisson[\exp(\mathbf{z}\boldsymbol{\delta}_2 + a_1)], \tag{1.2}$$

then the conditional density of y_2 is specified as:

$$f(y_2|\mathbf{z}, a_1) = \frac{[\exp(\mathbf{z}\delta_2 + a_1)]^{y_2} \exp\left[-\exp(\mathbf{z}\delta_2 + a_1)\right]}{y_2!},$$
(1.3)

where a_1 is assumed to be independent of \mathbf{z} , and $\exp(a_1)$ is distributed as $Gamma(\delta_0, 1/\delta_0)$ using a single parameter δ_0 , with $E(\exp(a_1)) = 1$ and $Var(\exp(a_1)) = 1/\delta_0$. The presence of a_1 in both equations (1.1) and (1.2) is what makes y_2 potentially endogenous in the equation of interest, (1.1). To illustrate this point, we could use, for example, u_2 instead of a_1 in the reduced form and u_1 instead of $\eta_1 a_1$ in the structural conditional mean and assume a linear function: $u_1 = \eta_1 u_2 + e_1$. Substitute the right-hand-side of this function into the structural conditional mean and then omit e_1 through multiplying all the coefficients by the scale factor $1/\sqrt{1 + \sigma_e^2}$, we will see $\eta_1 u_2$ and u_2 appear in the places of a_1 and $\eta_1 a_1$ in equations (1.1) and (1.2), respectively. Hence, rather than using u_1 and u_2 , we simply use a_1 and $\eta_1 a_1$ as stated in the reduced form and the structural conditional mean to govern the endogeneity of y_2 .

After a transformation (see Appendix D.4. for the derivation), the distribution of a_1 is derived as follows:

$$f(a_1; \delta_0) = \frac{\delta_0^{\delta_0} [\exp(a_1)]^{\delta_0} \exp(-\delta_0 \exp(a_1))}{\Gamma(\delta_0)}.$$
 (1.4)

In order to get the conditional mean $E(y_1|y_2, \mathbf{z})$, I specify the conditional density function of a_1 . Using Bayes' rule, it is:

$$f(a_1|y_2, \mathbf{z}) = \frac{f(y_2|a_1, \mathbf{z})f(a_1|\mathbf{z})}{f(y_2|\mathbf{z})}$$

Since $y_2|\mathbf{z}, a_1$ has a Poisson distribution and $\exp(a_1)$ has a gamma distribution, $y_2|\mathbf{z}$ is Negative Binomial II distributed, as a standard result (see the Poisson and Negative Binomial II models in Cameron & Trivedi (1986) and a specific derivation of this result from equations (D.1) to (D.3) in Appendix D).

After some algebra, the conditional density function of a_1 is:

$$f(a_1|y_2, \mathbf{z}) = \frac{\exp\left[\mathbf{P}\right] \left[\delta_0 + \exp(\mathbf{z}\delta_2)\right]^{(y_2 + \delta_0)}}{\Gamma(y_2 + \delta_0)},$$
(1.5)

where $\mathbf{P} = -\exp(\mathbf{z}\delta_2 + a_1) + a_1(y_2 + \delta_0) - \delta_0\exp(a_1)$.

The conditional mean $E(y_1|y_2, \mathbf{z})$ therefore will be obtained as:

$$E(y_1|y_2,\mathbf{z}) = \int_{-\infty}^{+\infty} \Phi(\alpha_1 y_2 + \mathbf{z}_1 \delta_1 + \eta_1 a_1) f(a_1|y_2,\mathbf{z}) da_1 = \mu(\theta; y_2, \mathbf{z}),$$
(1.6)

where $f(a_1|y_2, \mathbf{z})$ is given in (1.5) and $\theta = (\alpha_1, \delta_1, \eta_1)$.

The key to obtain the conditional mean of interest is to get the conditional density function of a_1 . Therefore, we need to assume the distribution of a_1 and specify $f(a_1|y_2, \mathbf{z})$ as above. For estimating purpose, it is necessary to compute $f(a_1|y_2, \mathbf{z})$ in (1.5) based on the parameters in the reduced form (1.2). These parameters can be estimated using a Negative Binomial II regression. And henceforth, they can be viewed as first-step estimated parameters. In the second step, we are interested in estimating conditional mean parameters, θ , in the FRMs.

For FRMs, we can consider a beta distribution or log-odds transformation of the fractional dependent variable. However, Wooldridge (2010) shows that these two approaches have some drawbacks. First, they rule out the case when the fractional response variable has some pileup at zero and/or one. Second, specifying a beta distribution is not robust and produces inconsistent estimators if any aspect of the distribution is misspecified. Third, the log-odds approach does not give a direct estimate of the conditional mean which is of interest; since this approach offers only the estimate of the transformed dependent variable (see more discussion in Papke & Wooldridge (1996)). Therefore, for the dependent variable which has some mass point at 0 and/or 1, and continuous in (0,1), we can focus on estimating the conditional mean of the fractional response (as stated in equation 1.1) that keeps the predicted value in the unit interval, and obtain robust estimators using the QMLE/NLS under the correctly specified conditional mean function. See Papke & Wooldridge (1996, 2008) for further details.

Given the fractional probit conditional mean model as in equation (1.1), there are many ways to estimate θ consistently. One possibility is to adopt the NLS estimator. This estimator is consistent and \sqrt{N} asymptotically normal. However, this estimator is unlikely to be asymptotically efficient because homoskedasticity is unlikely to hold for y_1 , even if we ignore the conditional Poisson distribution for y_2 . It might also be computationally intensive to obtain the weighting matrix for the NLS estimator. Hence, we can use a simpler, robust and efficient estimator, that is, the quasimaximum likelihood estimator (QMLE).

One can consider the QMLE using the Bernoulli distribution or the Poisson distribution of y_1 . The QMLE is simple and strongly consistent even if the true distribution of y_1 is not Bernoulli once the first moment is assumed to be correctly specified. There are other reasons that make the Bernoulli QMLE more attractive. First, maximizing the Bernoulli log likelihood is easy. Second, the Bernoulli distribution is a member of the linear exponential family (LEF) and it does not have any restriction as other distributions (see further discussion in Papke & Wooldridge (1996)). Moreover, it has some advantage over the Poisson distribution. For example, it is consistent with the nature of a fractional response variable which has both continuous and discrete characteristics. The Poisson distribution is consistent with a non-negative response variable but does not take into account mass points at 0 and/or 1. In addition, even though the Poisson distribution is a member of the LEF, it is chosen if we want the variance to be proportional to the mean, which is not realistic for a fractional response variable. It is unlikely that the variance is monotonically increasing in the mean. Another attraction of the Bernoulli QMLE is that it is efficient in a class of estimators containing all QMLEs in the LEF as long as the conditional mean is correctly specified and the variance assumption holds. The assumption that the variance associated with the quasi-log likelihood in equation (1.6) is the Bernoulli generalized linear models (GLM) variance will hold if the number of Bernoulli draws is independent of z_i . This assumption still holds in an empirical example of this chapter. However, in other applications, there is no guarantee that this assumption holds and it is recommended to obtain fully robust sandwich standard errors (see more discussion in Papke & Wooldridge (1996) and (Wooldridge, 2010, section 18.6)).

Therefore, in what follows, we use the QMLE or NLS with the Bernoulli quasi-log likelihood function to estimate θ of equation (1.6) in the second step.

The Bernoulli quasi-log likelihood function is given by:

$$l_i(\theta) = y_{1i} \ln \mu_i + (1 - y_{1i}) \ln(1 - \mu_i).$$
(1.7)

The QMLE of θ in the second step is obtained from the maximization problem (see more details in Appendix D.1.):

$$\max_{\theta \in \Theta} \sum_{i=1}^{n} l_i(\theta).$$
(1.8)

The NLS estimator of θ in the second step is attained from the minimization problem (see more

details in Appendix D.3.):

$$\min_{\theta \in \Theta} N^{-1} \sum_{i=1}^{N} [y_{1i} - \mu_i(\theta; y_{2i}, \mathbf{z}_i)]^2 / 2.$$
(1.9)

After we obtained the estimated parameters from the first-step and approximate the conditional mean (the detailed approximation procedure is discussed below), we estimate θ using the QMLE and NLS estimators as described in the above maximization and minimization problems. These estimators are the so-called two-step M-estimators that are consistent and asymptotically normal (see further discussion of these estimators in Newey & McFadden (1994) and (Wooldridge, 2002, chapter 12)).

Since $\mu_i = E(y_1|y_2, \mathbf{z})$ does not have a closed form solution, it is necessary to use a numerical approximation. The numerical routine for integrating out the unobserved heterogeneity in the conditional mean equation (1.6) is based on the Adaptive Gauss-Hermite quadrature. This adaptive approximation has proven to be more accurate with fewer points than the ordinary Gauss-Hermite approximation. The quadrature locations are shifted and scaled to be under the peak of the integrand. Therefore, the adaptive quadrature is performed well with an adequate amount of points (see more in Skrondal & Sophia (2004)).

Using the Adaptive Gauss-Hermite approximation, the above integral (1.6) can be obtained as:

$$\mu_{i} = \int_{-\infty}^{+\infty} h_{i}(y_{2i}, \mathbf{z}_{i}, a_{1}) da_{1} \approx \sqrt{2} \widehat{\sigma}_{i} \sum_{m=1}^{M} w_{m}^{*} \exp\left\{(a_{m}^{*})^{2}\right\} h_{i}(y_{2}, \mathbf{z}_{i}, \sqrt{2} \widehat{\sigma}_{i} a_{m}^{*} + \widehat{w}_{i}), \qquad (1.10)$$

where $\hat{\sigma}_i$ and \hat{w}_i are the adaptive parameters for observation *i*, w_m^* are the weights and a_m^* are the evaluation points, and *M* is the number of quadrature points. The approximation procedure follows Skrondal & Sophia (2004). The adaptive parameters $\hat{\sigma}_i$ and \hat{w}_i are updated in the k^{th} iteration of the optimization for μ_i with:

$$\mu_{i,k} \approx \sum_{m=1}^{M} \sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp\{(a_m^*)^2\} h_i(y_{2i}, \mathbf{z}_i, \sqrt{2} \hat{\sigma}_{i,k-1} a_m^* + \hat{\omega}_{i,k-1}),$$
$$\hat{\omega}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1}) \frac{\sqrt{2} \hat{\sigma}_{i,k-1} w_m^* \exp\{(a_m^*)^2\} h_i(y_{2i}, \mathbf{z}_i, \tau_{i,m,k-1})}{\mu_{i,k}},$$

$$\hat{\sigma}_{i,k} = \sum_{m=1}^{M} (\tau_{i,m,k-1})^2 \frac{\sqrt{2}\hat{\sigma}_{i,k-1} w_m^* \exp\{(a_m^*)^2\} h_i(y_{2i}, \mathbf{z}_i, \tau_{i,m,k-1})}{\mu_{i,k}} - (\hat{\omega}_{i,k})^2,$$

where

$$\tau_{i,m,k-1} = \sqrt{2}\hat{\sigma}_{i,k-1}a_m^* + \hat{\omega}_{i,k-1}.$$

This process is repeated until $\hat{\sigma}_{i,k}$ and $\hat{\omega}_{i,k}$ have converged for this iteration at observation *i* of the maximization algorithm. This adaptation is applied to every iteration until the log-likelihood difference from the last iteration is less than a relative difference of $1e^{-5}$; after this adaptation, the adaptive parameters are fixed.

Once the evaluation of the conditional mean has been done for all observations, the numerical values can be passed on to a maximizer in order to find the QMLE or NLS $\hat{\theta}$.

I summarize the method for estimating θ with the following procedure:

1.2.1 Estimation Procedure

- (i) Estimate δ_2 and δ_0 by using maximum likelihood of y_{i2} on \mathbf{z}_i in the Negative Binomial model. Obtain the estimated parameters $\hat{\delta}_2$ and $\hat{\delta}_0$.
- (ii) Use the fractional probit QMLE (or NLS) of y_{i1} on y_{i2} , z_{i1} to estimate α_1 , δ_1 and η_1 with the approximated conditional mean. The conditional mean is approximated using the estimated parameters in the first step and using the Adaptive Gauss-Hermite method.

After getting all the estimated parameters $\hat{\theta} = (\hat{\alpha}_1, \hat{\delta}_1, \hat{\eta}_1)'$, the standard errors in the second stage should be adjusted for the first stage estimation and obtained using the delta method. The standard errors obtained by using the delta method can be derived with the following formula:

$$\widehat{Avar}(\widehat{\widehat{\theta}}) = \frac{1}{N} \widehat{\mathbf{A}}_1^{-1} \left(N^{-1} \sum_{i=1}^N \widehat{\mathbf{r}}_{i1} \widehat{\mathbf{r}}_{i1}' \right) \widehat{\mathbf{A}}_1^{-1}.$$
(1.11)

For more details, see the derivation and matrix notation from equation (D.6) to equation (D.20) in Appendix D.1.

1.2.2 Average Partial Effects

Econometricians are often interested in estimating the average partial effects of explanatory variables in non-linear models in order to get comparable magnitudes with other nonlinear models and linear models. The Average Partial Effects (APE) can be obtained by taking the derivatives or the differences of a conditional mean equation with respect to the explanatory variables of interest. The APE cannot be estimated with the presence of unobserved factor. It is necessary to "integrate out" the unobserved variable in the conditional mean or average the partial effects across the distribution of the unobservable. Then we will obtain a single factor by taking the average across the sample in order to compare with the corresponding linear estimate.

I begin by reviewing the calculation of APEs when the explanatory variables are exogenous, following Papke & Wooldridge (2008), and then show how to identify the APEs with a count endogenous explanatory variable.

1.2.2.1 The Case with Exogenous Covariates and a Normally Distributed Heterogeneity

In a FRM with all exogenous covariates, model (1.1) with y_2 exogenous and a normally distributed a_1 is considered (for a general discussion of a FRM with all exogenous covariates, see Papke & Wooldridge (2008)).

Let $\mathbf{w} = (y_2, \mathbf{z}_1)$, dropping observation index *i*, equation (1.1) is rewritten as:

$$E(\mathbf{y}_1|\mathbf{w},a_1) = \mathbf{\Phi}(\mathbf{w}\boldsymbol{\beta} + a_1),$$

where **w** is the fixed terms and a_1 is the random term. We can also allow elements of **w** to be any function of (y_2, \mathbf{z}_1) , including nonlinear functions, such as quadratic or cubic forms, and interactions. If w_1 is continuous, then the partial effect with respect to w_1 is:

$$\partial E(y_1|\mathbf{w}, a_1) / \partial w_1 = \beta_1 \phi(\mathbf{w}\beta + a_1).$$

If w_1 is a dummy variable, we compute:

$$\Phi(\mathbf{w}^1\boldsymbol{\beta} + a_1) - \Phi(\mathbf{w}^0\boldsymbol{\beta} + a_1),$$

where \mathbf{w}^1 and \mathbf{w}^0 are two different values of the covariates including $w_1=1$ and $w_1=0$, respectively.

Since a_1 is not observed but we assume $a_1 | \mathbf{w} \sim Normal(0, \sigma_a^2)$, we can obtain the APE by averaging the partial effects across the distribution of a_1 :

$$E_{a_1}[\beta_1\phi(\mathbf{w}\beta+a_1)],$$

$$E_{a_1}[\Phi(\mathbf{w}^1\boldsymbol{\beta}+a_1)-\Phi(\mathbf{w}^0\boldsymbol{\beta}+a_1)],$$

and these are equivalent to getting: $\beta_1 \phi(\mathbf{w}\beta_a)$ and $[\Phi(\mathbf{w}^1\beta_a) - \Phi(\mathbf{w}^0\beta_a)]$ where subscript *a* stands

for division by $\sqrt{1 + \sigma_a^2}$.

Then we can obtain a single number to compare with the linear estimates by averaging the derivative or the difference across the sample.

For a continuous z_{11} , the APE is estimated by:

$$\hat{\delta}_{11a}\left[N^{-1}\sum_{i=1}^N\phi(\hat{\alpha}_{1a}y_{2i}+\mathbf{z}_{1i}\hat{\delta}_{1a})\right].$$

For a count variable y_2 , the APE is estimated by:

$$N^{-1}\sum_{i=1}^{N} \left[\Phi(\hat{\alpha}_{1a}y_{2}^{1} + \mathbf{z}_{1i}\hat{\delta}_{1a}) - \Phi(\hat{\alpha}_{1a}y_{2}^{0} + \mathbf{z}_{1i}\hat{\delta}_{1a}) \right].$$

For example, if we are interested in obtaining the APE when y_2 changes from 0 to 1, it is necessary to predict the difference in mean responses with $y_2 = 1$ and $y_2 = 0$ and average the difference across all units.

1.2.2.2 The Case with a Count Endogenous Covariate and a Non-normally Distributed Heterogeneity

In a fractional response model with a count endogenous variable, model (1.1) is considered with the estimation procedure provided in the previous section. The APEs are obtained by taking the derivatives or the differences in:

$$E_{a_1}[\Phi(\alpha_1 y_2^0 + \mathbf{z}_1^0 \delta_1 + \eta_1 a_1)], \qquad (1.12)$$

with respect to the elements of (y_2^0, \mathbf{z}_1^0) . In the argument of the expectations operator, (y_2^0, \mathbf{z}_1^0) are fixed terms and a_1 is a random term.

The partial effect (PE) is obtained for a continuous variable z_{11} :

$$PE(y_2^0, \mathbf{z}_1^0, a_1) = \delta_{11}\phi(\alpha_1 y_2^0 + \mathbf{z}_1^0 \delta_1 + \eta_1 a_1),$$
(1.13)

and for a discrete variable y_2 , we compute:

$$\Phi(\alpha_1 y_2^1 + \mathbf{z}_1^0 \delta_1 + \eta_1 a_1) - \Phi(\alpha_1 y_2^0 + \mathbf{z}_1^0 \delta_1 + \eta_1 a_1),$$
(1.14)

which is the difference in mean responses with two fixed points: $y_2 = y_2^1$ and $y_2 = y_2^0$ that we are interested in.

To obtain the APEs, we need to average the above partial effects across the distribution of a_1 :

$$APE_{c} = E_{a_{1}}[\delta_{11}\phi(\alpha_{1}y_{2}^{0} + \mathbf{z}_{1}^{0}\delta_{1} + \eta_{1}a_{1})], \qquad (1.15)$$

for the continuous case, and

$$APE_{d} = E_{a_{1}}[\Phi(\alpha_{1}y_{2}^{1} + \mathbf{z}_{1}^{0}\delta_{1} + \eta_{1}a_{1}) - \Phi(\alpha_{1}y_{2}^{0} + \mathbf{z}_{1}^{0}\delta_{1} + \eta_{1}a_{1})], \qquad (1.16)$$

for the discrete case.

This is equivalent to integrate out a_1 and we respectively receive:

$$\psi = APE_c = \int_{-\infty}^{+\infty} \delta_{11}\phi(\alpha_1 y_2^0 + \mathbf{z}_1^0 \delta_1 + \eta_1 q_1) f(q_1 | y_2^0, \mathbf{z}_1^0; \theta) dq_1,$$
(1.17)

$$\lambda = APE_d = \int_{-\infty}^{+\infty} \Phi(\mathbf{g}^1 \theta) f(q_1 | y_2^1, \mathbf{z}_1^0; \theta) dq_1 - \int_{-\infty}^{+\infty} \Phi(\mathbf{g}^0 \theta) f(q_1 | y_2^0, \mathbf{z}_1^0; \theta) dq_1, \quad (1.18)$$

where q_1 is a dummy argument in the integration, $\mathbf{g}^1 = (y_2^1, \mathbf{z}_1^0, q_1)$, and $\mathbf{g}^0 = (y_2^0, \mathbf{z}_1^0, q_1)$.

These APEs are estimated by:

$$\widehat{APE}_{c} = \widehat{\delta}_{11} \int_{-\infty}^{+\infty} \phi(\widehat{\alpha}_{1}y_{2}^{0} + \mathbf{z}_{1}^{0}\widehat{\delta}_{1} + \widehat{\eta}_{1}q_{1})f(q_{1}|y_{2}^{0}, \mathbf{z}_{1}^{0}; \widehat{\theta})dq_{1},$$
(1.19)

$$\widehat{APE}_{d} = \int_{-\infty}^{+\infty} \Phi(\mathbf{g}^{1}\hat{\theta}) f(q_{1}|y_{2}^{1}, \mathbf{z}_{1}^{0}; \hat{\theta}) dq_{1} - \int_{-\infty}^{+\infty} \Phi(\mathbf{g}^{0}\hat{\theta}) f(q_{1}|y_{2}^{0}, \mathbf{z}_{1}^{0}; \hat{\theta}) dq_{1},$$
(1.20)

Since equations (1.19) and (1.20) cannot be obtained in a closed form, we need to use the Adaptive Gauss-Hermite method to approximate the density of $f(q_1|y_2^k, \mathbf{z}_1^0; \theta)$; k = 0, 1. This is equivalent to obtain:

$$\hat{\boldsymbol{\psi}} = \widehat{APE}_c \approx \hat{\boldsymbol{\delta}}_{11} \left\{ \sqrt{2} \widehat{\sigma}_i \sum_{m=1}^M \{ w_m^* \exp[(a_m^*)^2] \} \boldsymbol{\phi}(\mathbf{g}^0 \widehat{\boldsymbol{\theta}}) f(y_2^0, \mathbf{z}_1^0, (\sqrt{2} \widehat{\sigma}_i a_m^* + \widehat{w}_i); \widehat{\boldsymbol{\theta}}) \right\}, \quad (1.21)$$

$$\widehat{\lambda} = \widehat{APE}_d = \widehat{\lambda}^1 - \widehat{\lambda}^0, \qquad (1.22)$$

where

$$\widehat{\lambda}^{k} = \sqrt{2}\widehat{\sigma}_{i}\sum_{m=1}^{M} \{w_{m}^{*}\exp[(a_{m}^{*})^{2}]\} \Phi(\mathbf{g}^{k}\widehat{\theta})f(y_{2}^{k},\mathbf{z}_{1}^{0},(\sqrt{2}\widehat{\sigma}_{i}a_{m}^{*}+\widehat{w}_{i});\widehat{\theta}),$$
(1.23)

in addition to $\mathbf{g}^k = (y_2^k, \mathbf{z}_1^0, (\sqrt{2}\widehat{\sigma}_i a_m^* + \widehat{w}_i)); k = 0, 1$ and $\theta = (\alpha_1, \delta_1, \eta_1)'$. For a comparison between the linear model estimates and the fractional probit estimates, it is useful to have a single factor. This single factor can be obtained by averaging out \mathbf{z}_{1i} across all individuals in the formula of $\hat{\psi}$ and $\hat{\lambda}$. For example, in order to get the APE when y_2 changes from 0 to 1, it is necessary to predict the difference in the mean responses with $y_2 = 0$ and $y_2 = 1$ and take the average of the differences across all units. This APE gives us a number comparable to the linear model's estimate.

The standard errors for the APEs will be obtained using the delta method. The detailed derivation is provided from equation (D.21) to equation (D.40) in Appendix D.1.

1.3 Monte Carlo Simulations

This section examines the finite sample properties of the two-step QML and NLS estimators of the population averaged partial effect in a fractional response model with a count endogenous variable. Some Monte Carlo experiments are conducted to compare these estimators with other estimators under different scenarios. These estimators are evaluated under correct model specification with different degrees of endogeneity; with strong and weak instrumental variables; and with different sample sizes. The behavior of these estimators is also examined with respect to a choice of a particular distributional assumption.

1.3.1 Estimators

Two sets of estimators under two corresponding assumptions are considered: (1) y_2 is assumed to be exogenous, and (2) y_2 is assumed to be endogenous. Under the former assumption, three estimators are used: the ordinary least squares (OLS) estimator in a linear model, the maximum likelihood estimator (MLE) in a Tobit model and the quasi-maximum likelihood estimator (QMLE) in a fractional probit model. Under the latter assumption, five estimators are examined: the twostage least squares (2SLS) estimator, the two-step maximum likelihood estimator (MLE) in a Tobit model using the Blundell-Smith estimation method (hereafter the Tobit BS), the two-step QMLE in a fractional probit model using the Papke-Wooldridge's estimation method (hereafter the QMLE-PW; see more discussion of handling endogeneity in Papke & Wooldridge (2008)), the two-step QMLE and the two-step NLS estimators in a fractional probit model using the estimation method proposed in the previous section.

1.3.2 Data Generating Process

The count endogenous variable is generated from a conditional Poisson distribution:

$$f(y_{2i}|x_{1i}, x_{2i}, z_i, a_{1i}) = \frac{\exp(-\lambda_i)\lambda_i^{y_{2i}}}{y_{2i}!},$$
(1.24)

11-

with a conditional mean:

$$\lambda_i = E(y_{2i}|x_{1i}, x_{2i}, z_i, a_{1i}) = \exp(\delta_{21}x_{1i} + \delta_{22}x_{2i} + \delta_{23}z_i + \rho_1a_{1i}),$$
(1.25)

using independent draws from normal distributions: $z \sim N(0, 0.3^2)$, $x_1 \sim N(0, 0.2^2)$, $x_2 \sim N(0, 0.2^2)$ and $\exp(a_1) \sim Gamma(1, 1/\delta_0)$ where 1 and $1/\delta_0$ are the mean and variance of a gamma distribution. Parameters in the conditional mean model are set to be:

$$(\delta_{21}, \delta_{22}, \delta_{23}, \rho_1, \delta_0) = (0.01, 0.01, 1.5, 1, 3).$$

The dependent variable is generated by first drawing a binomial random variable *x* with *n* trials and a probability *p* and then $y_1 = x/n$. In this simulation, n = 100 and *p* comes from a conditional normal distribution with the conditional mean:

$$p = E(y_{1i}|y_{2i}, x_{1i}, x_{2i}, a_{1i}) = \Phi(\delta_{11}x_{1i} + \delta_{12}x_{2i} + \alpha_1y_{2i} + \eta_1a_{1i}),$$
(1.26)

and parameters in this conditional mean are set at: $(\delta_{11}, \delta_{12}, \alpha_1, \eta_1) = (0.1, 0.1, -1.0, 0.5).$

In order to compare the magnitudes between a nonlinear model and a linear model, we are interested in computing APEs. Based on the population values of the parameters set above, the so-called true value of the APE with respect to each variable is reported as the mean of the APEs approximated via the simulations with the standard procedure described below.

First, when y_2 is treated as a continuous variable, the so-called true value of the APE with respect to y_2 is approximated from simulations by first computing the derivative of the conditional mean with respect to y_2 , and then taking the average across the distribution of a_1 :

$$APE = -1.0 * \frac{1}{N} \sum_{i=1}^{N} \phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * y_2 + 0.5 * a_{1i}).$$
(1.27)

Now when y_2 is allowed to be a count variable, the so-called true values of the APEs with respect to y_2 are computed by first taking differences in the conditional mean. These true values of the APEs are computed at interesting values. In this chapter, I will take the first three examples when y_2 increases from 0 to 1, 1 to 2 and 2 to 3, respectively and the corresponding true values of the APEs are:

$$APE01 = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 1 + 0.5 * a_{1i}) \\ -\Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 0 + 0.5 * a_{1i}) \end{bmatrix},$$
(1.28)

$$APE12 = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 2 + 0.5 * a_{1i}) \\ -\Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 1 + 0.5 * a_{1i}) \end{bmatrix},$$
(1.29)

$$APE23 = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} \Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 3 + 0.5 * a_{1i}) \\ -\Phi(0.1 * x_1 + 0.1 * x_2 - 1.0 * 2 + 0.5 * a_{1i}) \end{bmatrix}.$$
 (1.30)

These so-called true values of the APEs (which are approximated through simulations) with respect to y_2 and other exogenous variables are reported in Tables A.1-A.4. The experiment is conducted with 500 replications and the sample size is normally set at 1000 observations.

1.3.3 Experiment Results

I report sample means, sample standard deviations (SD) and root mean squared errors (RMSE) of these 500 estimates. In order to compare estimators across linear and non-linear models, I am interested in comparing the APE estimates from different models.

1.3.3.1 Simulation Result with a Strong Instrumental Variable

Tables A.1-A.4 report the simulation outcomes of the APE estimates for the sample size N = 1000 with a strong instrumental variable (IV) and different degrees of endogeneity, where $\eta_1 = 0.1$, $\eta_1 = 0.5$, and $\eta_1 = 0.9$. The IV is strong in the sense that the coefficient on *z* is $\delta_{23} = 1.5$ in the first stage and the F-statistics on the significance test of *z* in the first-stage are large. These F-statistics have average values equivalent to 91.75, 107.57 and 133.56 in 500 replications for three designs of η_1 : $\eta_1 = 0.1$, $\eta_1 = 0.5$, and $\eta_1 = 0.9$, respectively. Three different values of η_1 are selected which corresponds to low, medium and high degrees of endogeneity. Columns 2-10 contain the true values of the APE estimates and the means, SD and RMSE of the APE estimates from different models with different estimation methods. Columns 3-5 consist the means, SD and RMSE of the APE estimates for all variables from 500 replications with y_2 assumed to be exogenous. Columns 6-10 include the means, SD and RMSE of the APE estimates from 500 replications with y_2 allowed to be endogenous.

I first report the simulation outcomes for the sample size N = 1000 and $\eta_1 = 0.5$ (see Table A.1). The APE estimates using the proposed methods of QMLE and NLS in columns 9-10 are closest to the true values of the APEs when y_2 is discrete (-.3200, -.1273 and -.0212). It is typical to get these three APEs for a discrete y_2 (when y_2 goes from 0 to 1, 1 to 2 and 2 to 3) as examples in order to see the pattern of the means, SD and RMSE of the APE estimates. The APE estimate is also very close to the true value of the APE (-.2347) when y_2 is treated as a continuous variable. Table A.1 shows that the OLS estimate is about a half of the true value of the APE. The first source of large bias in the OLS estimate comes from the ignorance of the endogeneity in the count variable y_2 (with $\eta_1 = 0.5$). The second source of bias in the OLS estimate is due

to the neglect of the non-linearity in both the structural and reduced-form equations (1.1) and (1.2). The 2SLS approach also produces a biased estimator of the APE because of the second reason mentioned above even though the endogeneity is taken into account. The MLE estimators in the Tobit model have smaller bias than the estimators in the linear model but larger bias than the estimators in the fractional probit model because they do not consider the functional form of the fractional response variable and the count explanatory variable. When the endogeneity is corrected, the MLE estimator in the Tobit model using Blundell-Smith method has a smaller bias than the counterpart where y_2 is assumed to be exogenous. Among the fractional probit models, the two-step QMLE estimator, where y_2 is assumed to be exogenous, (column 5) has the largest bias because it ignores the endogeneity of y_2 . However, it still has a smaller bias than other estimators of the linear and Tobit models. The two-step QMLE-PW estimator (column 8) provides useful result because its estimates are also very close to the true values of the APEs but it produces a larger bias than the two-step QMLE and NLS estimators proposed in this chapter. Similar to the two-step MLE estimator in Tobit model using Blundell-Smith method, the two-step QMLE-PW estimator adopts the control function approach. This approach utilizes the linearity in the first stage equation. As a result, it ignores the discreteness in y_2 which leads to the larger bias than the two-step QMLE and NLS estimators proposed in this chapter.

The first set of estimators with y_2 assumed to be exogenous (columns 3-5) has relatively smaller SDs than the second set of estimators with y_2 allowed to be endogenous (columns 6-10) because the methods that correct for endogeneity using IVs have more sampling variation than their counterparts without endogeneity correction. This results from the less-than-unit correlation between the instrument and the endogenous variable. However, the SDs of the two-step QMLE and NLS estimators (columns 9-10) are no worse than the QMLE estimator where y_2 assumed to be exogenous (column 5).

Among all estimators, the two-step QMLE and NLS estimators proposed in this chapter have the smallest RMSE, not only for the case where y_2 is allowed to be a discrete variable but also for the case where y_2 is treated as a continuous variable using the correct model specification. As discussed previously, the two-step QMLE estimator using Papke-Wooldridge method has the third smallest RMSE since it also uses the same fractional probit model. Comparing columns 3 and 6, 4 and 7, 5 and a set of all columns 8-10, the RMSEs of the methods correcting for endogeneity are smaller than those of their counterparts.

Table A.2 reports simulation result for coefficient estimates. The coefficient estimates are useful in the sense that it gives the directions of the effects. For studies which only require exploring the signs of the effects, the coefficient tables are necessary. For studies which require comparing the magnitudes of the effects, we essentially want to estimate the APEs. Table A.2 shows that the means of point estimates are close to their true values for all parameters using the two-step QML (or NLS) approach (-1.0, 0.1 and 0.1). The bias is large for both 2SLS method and OLS method. These results are as expected because the 2SLS method uses the predicted value from the first stage OLS so it ignores the distributional information of the right-hand-side (RHS) count variable, regardless of the functional form of the fractional response variable. The OLS estimates do not carry the information of endogeneity. Both the 2SLS and OLS estimates are biased because they do not take into account the presence of unobserved heterogeneity. The bias for a Tobit Blundell-Smith model is similar to the bias with the 2SLS method because it does not take into account the distributional information of the right-hand-side count variable and it employs a different functional form given the fact that the fractional response variable has a small number of zeros. The biases for both the QMLE estimator treating y_2 as an exogenous variable and for the two-step QMLE-PW estimator are larger than those of the two-step QMLE and NLS estimators in this chapter. In short, simulation results indicate that the means of point estimates are close to their true values for all parameters using the two-step QMLE and the NLS approach mentioned in the previous section.

Simulations with different degrees of endogeneity through the coefficient $\eta_1 = 0.1$ and $\eta_1 = 0.9$ are also conducted (see Table A.3 and A.4). Not surprisingly, with less endogeneity, $\eta_1 = 0.1$, the set of the estimators treating y_2 as an exogenous variable produces the APE estimates closer to the true values of the APE estimates; the set of the estimators treating y_2 as an endogenous variable has the APE estimates further from the true values of the APE estimates. With more endogeneity,

 $\eta_1 = 0.9$, the set of the estimators treating y_2 as an endogenous variable has the APE estimates getting closer to the true values of the APE estimates; and the set of the estimators treating y_2 as an exogenous variable gives the APE estimates further from the true values of the APE estimates. As an example, it is noted that, as η_1 increases, the APE estimates of the 2SLS method are less biased while the APE estimates of the QMLE estimator treating y_2 as an exogenous variable are more biased and the difference between these two APE estimates is smaller since the endogeneity is corrected.

All other previous discussions on the bias, SD and RMSE still hold with $\eta_1 = 0.1$ and $\eta_1 = 0.9$. It confirms that the two-step QMLE and NLS estimators perform very well under different degrees of endogeneity.

1.3.3.2 Simulation Result with a Weak Instrumental Variable

Table A.5 reports the simulation outcomes of the APE estimates for the sample size N = 1000 with a weak IV and $\eta_1 = 0.5$. Using the rule of thumb on a weak instrument (suggested in Staiger & Stock (1997)), the coefficient on *z* is chosen as $\delta_{23} = 0.3$ which corresponds to a very small firststage F-statistic (the mean of the F-statistic is 6.97 in 500 replications). Columns 2-10 contain the true values of the APE estimates, the means, SD and RMSE of the APE estimates from different models with different estimation methods. Columns 3-5 consist the means, SD and RMSE of the APE estimates for all variables from 500 replications with y_2 assumed to be exogenous. Columns 6-10 include the means, SD and RMSE of the APE estimates for all variables from 500 replications with y_2 allowed to be endogenous. The simulation results show that, even though the instrument is weak, the set of estimators assuming y_2 endogenous still has smaller bias than the set of estimators assuming y_2 exogenous. The two-step QMLE and NLS APE estimates are still very close to the true values of the APEs for both cases in which y_2 is treated to be a continuous variable and y_2 is allowed to be a count variable. Their SD and RMSE are still the lowest among the estimators considering y_2 endogenous. Table A.11 also provides this evidence.

Simulation results from my proposed procedure show that the two-step QMLE and NLS APE

estimates are less biased and more efficient compared with the linear model's and other models' estimates. However, at the first glance, we can notice the standard deviation in Table A.5 is less than the standard deviation from Table A.1 under columns of QMLE and NLS, which is contrary to the pattern of standard deviations in the linear model (under the column with 2SLS estimation method). The standard deviation from my proposed procedure (under the columns of QMLE and NLS) is smaller in the case of a weak IV than the case of a strong IV, which seems odd at first if we judge that exclusion restrictions are driving identification. If we look at the result from the column of 2SLS, the bias and inefficiency of 2SLS estimates may arise because a linear model may provide a poor approximation for the count and fractional response variable. It suggests that nonlinearities have larger contributions than exclusion restriction to identification in the nonlinear models. In other words, functional form assumptions are mainly responsible for identification rather than the exclusion restriction. Therefore, it is worth investigating the reason why the estimates in my proposed procedure tend to be closer to the true value of the APEs and are always efficient without increasing the standard deviation in the case of weak instrument. We design the experiment similar to simulated experiment in Table A.5 but the coefficient on the instrument is 0 ($\delta_{23} = 0$). This is equivalent to the case of no instruments. The results in Table A.6 show that standard deviations under columns of QMLE and NLS are still smaller suggesting nonlinearity is responsible for identification since there is no exclusion restriction here.

1.3.3.3 Simulation Result with Different Sample Sizes

Four sample sizes are chosen to represent those commonly encountered sizes in applied research. These range from small to large sample sizes: 100, 500, 1000 and 2000. Tables A.7-A.10 report the simulation outcomes of the APE estimates with a strong IV, $\eta_1 = 0.5$, for sample sizes N = 100, 500, 1000, and 2000 respectively. Table A.8 is equivalent to Table A.1. Columns 2-10 contain the true values of the APE estimates and the means, SD and RMSE of the APE estimates from different models with different estimation methods. Columns 3-5 consist the means, SD and RMSE of the APE estimates for the APE estimates from 500 replications with y_2 assumed to be exogenous. Columns 6-10 include the means, SD and RMSE of the APE estimates for all variables from 500 replications with y_2 allowed to be endogenous. In general, the simulation results indicate that the SD and RMSE for all estimators are smaller for larger sample sizes. Previous discussion as in 3.3.1 is still applied. The two-step QMLE and NLS estimators perform very well in all sample sizes with the smallest SD and RMSE. They are also the least biased estimators among all the estimators in this discussion.

1.3.3.4 Simulation Result with a Misspecifie Distribution

The original assumption is that $\exp(a_1) \sim Gamma(1, 1/\delta_0)$. However, misspecification is dealt with in this part. The distribution of $\exp(a_1)$ is no longer gamma, instead, $a_1 \sim N(0, 0.1^2)$ is assumed. The finite sample behavior of all the estimators in this incorrect specification is examined. Table A.11 shows the simulation results for the sample size N = 1000 with a strong IV and $\eta_1 = 0.5$ under misspecification. All of the previous discussion under the correct specification as in 1.3.3.1 is not affected. The APE estimates under the fractional probit model are still very close to the true values of the APEs.

Table A.12 shows the simulation results of the APE estimates with the sample size N = 1000and $\eta_1 = 0.5$. The estimates are close to true values of the APEs, with very small MSE and rejection rates close to 0.05.

We should note from all tables in the section of Monte Carlo simulations that the standard deviations under the columns of QMLE and NLS using the proposed procedure are not directly comparable as standard errors. However, we can see from the simulation result of Table A.13 that the proposed procedure's analytical variance is quite reliable because estimates of mean of standard errors using analytical computation is quite close to those using bootstrapping method.

1.3.4 Conclusion from the Monte Carlo Simulations

This section examined the finite sample behavior of the estimators proposed in the FRM with an endogenous count variable. The results of some Monte Carlo experiment show that the twostep QMLE and NLS estimators have smallest standard deviations, RMSE and least biased when the endogeneity is presented. The two-step QMLE and NLS methods also produce least biased estimates in terms of both parameters and the APEs compared to other alternative methods.

1.4 Application and Estimation Results

My proposed estimators can be applied in a model of female labor supply. The dependent variable refers to the allocation of total hours per week mothers spent on working. Hereafter, we name the dependent variable as weekly fractional working hours. The data in this chapter were used in Angrist & Evans (1998) to illustrate a linear model with a dummy endogenous variable: more than two kids. They estimate the effect of additional children on female labor supply, considering the number of children as endogenous and using the instruments: same sex and twins at the first two births. They found that married women who have the third child reduce their labor supply, and their 2SLS estimates are roughly a half smaller than the corresponding OLS estimates.

In this application, the fractional response variable is the fraction of total weekly hours that a woman spends working. This variable is generated from the number of working hours, which was used in Angrist & Evans (1998), divided by the maximum hours per week (168). There is a substantial number of women who do not spend any hours working, 13068 observations at zero. Therefore, a Tobit model might be a choice.

In this application, we are interested in estimating a model of weekly fractional working hours (FrHour) for women who take into consideration of having the number of children as a count endogenous factor. We begin with the linear model as follows:

$$FrHour = \alpha_1 Kidno + \delta_1 Educ + \delta_2 Age + \delta_3 Age fb + \delta_4 Hispan + \delta_5 NmInc + a_1.$$
(1.31)

The count variable in this application is the number of children beyond two, between 0 and 10, instead of an indicator for having more than two kids which was used in Angrist & Evans (1998). The number of kids is considered endogenous, which is in line with the recent existing empirical literature. First, the number and timing of children born are controlled by a mother makes fertility

decisions correlated with the number of children. Second, some women have preference for familybased activity or market-based work, so fertility is correlated with women's heterogeneity. The estimation sample contains 31,824 women, more than 50% is childless, 31% have one kid, 11% have two kids and the rest have more than two kids. Table A.14 gives the frequency distribution of the number of children and it appears to have excess zeros and long tails with the average number of children is around one. Other explanatory variables which are exogenous, including demographic and economic variables of the family, are also described in Table A.15.

The current research on parent's preferences over the sex mixture of their children using US data shows that most families would prefer at least one child of each sex. For example, Ben-Porath & Welch (1976) found that 56% of families with either two boys or two girls has a third birth while only 51% families with one boy and one girl had a third child. Angrist & Evans (1998) found that only 31.2% of women with one boy and one girl have a third child whereas 38.8% and 36.5% of women with two girls and two boys have a third child, respectively. With the evidence that women with children of the same sex are more likely to have additional children, the instruments that we can use are same sex and twins. Table A.16 illustrates the result of first-stage estimates with the significant statistics of same sex and twins.

Table A.17 shows the estimation results of the OLS in a linear model, the MLE in a Tobit model and the QMLE in a fractional probit model when y_2 is assumed exogenous. The estimation results of the 2SLS in a linear model, the MLE in a Tobit BS model, the QMLE-PW, the QMLE and NLS estimation in a fractional probit model are shown in Table A.18 when y_2 is assumed endogenous.

Since I also analyze the model using the Tobit BS model, its model specification and derivation of the conditional mean, the average partial effects and the estimation approach are included in Appendix D.2. The two-step NLS method with the same conditional mean used in the two-step QMLE method is also presented in Appendix D.3.

Ordinary least squares

The OLS estimation often plays a role as a benchmark since its computation is simple, its interpretation is straightforward and it requires fewer assumptions for consistency. The estimates

of a linear model in which the fraction of total working hours per week is the response variable and the number of kids is considered exogenous are provided in Table A.17. As discussed in the literature of women's labor supply, the coefficient of the number of kids is negative and statistically significant. The linear model with the OLS estimation ignores functional form issues that arise from the excess-zeros nature of the dependent variable. In addition, the predicted value of the fraction of the total weekly working hours for women always lies in the unit interval. The use of the linear model with the OLS estimation will not make any sense if the predicted value occurs outside this interval.

A Tobit model with an exogenous number of kids

There are two reasons that a Tobit model might be practical. First, the fraction of working hours per week has many zeros. Second, the predicted value needs to be nonnegative. The estimates are given in Table A.17. The Tobit coefficients have the same signs as the corresponding OLS estimates, and the statistical significance of the estimates is similar. For magnitude, the Tobit partial effects are computed to make them comparable to the linear model estimates. First of all, the partial effect of a discrete explanatory variable is obtained by estimating the Tobit conditional mean. Second, the differences in the conditional mean at two values of the explanatory variable that are of interest is computed (for example, we should first plug in $y_{2i} = 1$ and then $y_{2i} = 0$). As implied by the coefficient, having the first child reduces the estimated fraction of total weekly working hours by about 0.023, or 2.3 percentage points, a larger effect than 1.9 percentage points of the OLS estimate. Having the second child and the third child make the mother work less by about 0.021 or 2.1 percentage points and 0.018 or 1.8 percentage points, respectively. All of the OLS and Tobit statistics are fully robust and statistically significant. Comparing with the OLS partial effect, which is about 0.019 or 1.9 percentage points, the Tobit partial effects are larger for having the first kid but almost the same for the second and the third kid. The partial effects of continuous explanatory variables can be obtained by taking the derivatives of the conditional mean; or we can practically get the adjustment factors to make the adjusted Tobit coefficients roughly comparable to the OLS estimates. All of the Tobit coefficients given in Table A.17 for continuous variables are larger than the corresponding OLS coefficients in absolute values. However, the Tobit partial effects for continuous variables are slightly larger than the corresponding OLS estimates in absolute values.

A Fractional response model (FRM) with an exogenous number of kids

Following Papke & Wooldridge (1996), I also use the fractional probit model assuming the number of children exogenous for a comparison purpose. The FRM's estimates are similar to the Tobit's estimates, but they are even closer to the OLS estimates. The statistical significance of QML estimates is almost the same as that of the OLS estimates (see Table A.17). Having the second child reduces the estimated fraction of total weekly working hours by 1.9 percentage points, which is roughly the same as the OLS estimate. However, having the first and third child result in different partial effects. Having the first kid makes a mother work much less by 2.0 percentage points, and having the third kid makes a mother work less by 1.6 percentage points.

Two-stage least squares

In the literature on female labor supply, Angrist & Evans (1998) consider fertility endogenous. Their remarkable contribution is to use two binary instruments: genders of the first two births are the same (*samesex*) and twins at the first two births (*multi2nd*) to account for an endogenous third child. The 2SLS estimates are replicated and reported in Table A.18. The first stage estimates using the OLS method and assuming a continuous number of children, given in Table A.16, show that women with higher education are estimated to be 6.5 percentage points less likely to have kids. In magnitude, the 2SLS estimates are less than the OLS estimates for the number of kids but roughly the same for other explanatory variables. With IV estimates, having children leads a mother to work less by about 1.6 percentage points, which is smaller than the corresponding OLS estimates of about 1.9 percentage points. These findings are consistent with Angrist and Evans' result.

A Tobit BS model with an endogenous number of kids
A Tobit BS model is used with the number of children endogenous (see Table A.18). Only the Tobit average partial effect of the number of kids have statistically slightly larger effect than that of the 2SLS estimates. The APEs of the Tobit estimates are almost the same as those of the corresponding 2SLS estimates for other explanatory variables. Having the first, second and third kid reduce the fraction of hours a mother spends working per week by around 1.8, 1.7 and 1.5 percentage points, respectively. Having the third kid reduces a mother's fraction of working hours per week by the same amount as the 2SLS estimates. The statistical significance is almost the same for the number of kids. The Tobit BS method is similar to the 2SLS method in the sense that the first stage uses a linear estimation and it ignores the discrete nature of the number of children. It explains why the Tobit BS result gets very close to the 2SLS estimates.

A FRM with an endogenous number of kids

Now let us consider the FRM with the number of kids endogenous. The fractional probit model with Papke-Wooldridge method (2008) has dealt with the problem of endogeneity. However, this method has not taken into account the problem of count endogeneity. The endogenous variable in this model is treated as a continuous variable, hence, the partial effects at discrete values of the count endogenous variable are not considered. In this chapter, the APEs of the QMLE-PW estimates are also computed in order to be comparable with other APE estimates. Having the first kid reduces a mother's fraction of weekly working hours by the same amount as the 2SLS estimates. Treating the number of children continuous also gives the same effect as the 2SLS estimate on the number of kids. As the number of children increases, the more working hours a mother has to sacrifice. Having the second and third kids reduce the fraction of hours a mother spends working per week by around 1.6, 1.5 and 1.4 percentage points, respectively. The statistical significance is the same as the Tobit BS estimates for the number of kids. The APEs of the two-step QMLE-PW estimates are almost the same as those of the corresponding 2SLS estimates for other explanatory variables.

The fractional probit model with the methods proposed in this chapter is attractive because it controls for endogeneity, functional form issues and the presence of unobserved heterogeneity. More importantly, the number of children is considered a count variable instead of a continuous variable. Both the two-step QMLE and NLS are considered and the two-step NLS estimates are quite the same as the two-step QML estimates. The two-step QML and NLS's coefficients and robust standard errors are given in Table A.18 and the first-stage estimates are reported in Table A.16.

In the first stage, the Poisson model for the count variable is preferred because of two reasons. First, the distribution of the count variable with a long tail and excess zeros suggests an appropriate model of gamma heterogeneity instead of normal heterogeneity. Second, adding the unobserved heterogeneity with the standard exponential gamma distribution to the Poisson model transforms the model to the Negative Binomial model, which can be estimated by the maximum likelihood method. The OLS and Poisson estimates are not directly comparable. For instance, increasing education by one year reduces the number of kids by 0.065 as in the linear coefficient and by 7.8% as in the Poisson coefficient.

The fractional probit's estimates have the same signs as the corresponding OLS and 2SLS estimates. In addition, the result shows that the two-step QMLE is more efficient than the OLS and 2SLS estimators. For magnitude, the fractional probit's APEs are computed to make them comparable to the linear model's estimates. Similar to the Tobit model, the partial effect of a discrete explanatory variable is obtained by estimating the conditional mean and taking the differences at the values we are interested in. Regarding the number of kids, having more kids reduces the fraction of hours that a mother works weekly. Having the first child cuts the estimated fraction of total weekly working hours by about 0.017, or 1.7 percentage points, which is similar to the 2SLS estimates, and less than the OLS estimates. Having the second child and the third child make a mother work less by about 1.5 percentage points and 1.4 percentage points, respectively. Even though having the third kid reduces a mother's fraction of weekly working hours compared to having the second kid, the marginal reduction is less, since a marginal reduction of 0.2 percentage points for having the second kid now goes down to 0.1 percentage points for having the third kid. This can be seen as the "adaptation effect" as the mother adapts and works more effectively after having the first kid. The partial effects of continuous explanatory variables can be obtained by taking the derivatives of the conditional mean so that they would be comparable to the OLS, 2SLS estimates and other alternative estimates.

All of the estimates in Table A.18 tell a consistent story about fertility. Statistically, having any children reduces significantly a mother's working hours per week. In addition, the more kids a woman has, the more hours that she needs to forgo. The FRM treating the number of kids as endogenous and as a count variable gives an evidence that the marginal reduction of women's working hours per week is less as women have additional kids. In addition, the FRM's estimates, taking into account the endogeneity and count nature of the number of children, are statistically significant and more significant than the corresponding linear model's and Tobit BS's estimates.

One advantage of the fractional probit model with the two-step QMLE (NLS) method that we are discussing in this part is that it fits the data better than alternative models or methods. Either R-squared $(S_1 = SSE/SST = 1 - \left[\sum_{i=1}^n (y_{1i} - \hat{y}_1)^2 / \sum_{i=1}^n (y_{1i} - \bar{y}_1)^2\right])$ or the correlation squared $(S_2 = \{Corr[y_1, E(y_1|y_2, \mathbf{z})]\}^2)$ can be used to compare the goodness-of-fit among these models. The statistics on fractional probit-QMLE, NLS, Tobit BS, and Linear 2SLS are 0.116, 0.114, 0.090, and 0.088, respectively. This shows that the fractional probit model using the two-step QMLE(NLS) methods has larger goodness-of-fit statistic(s) than that of the Tobit model using Blundell-Smith's procedure and the linear model using the 2SLS method.

It seems questionable that the standard errors under the columns of QMLE and NLS methods (see Table A.18) are unexpectedly smaller than alternative methods' standard errors if we compare with the simulation results in Table A.1. There are two things we need to make clear: i) In the simulation results, they are standard deviations instead of standard error estimates; and ii) Table A.1 show the case of a strong IV whereas our empirical results does not have a strong IV. Therefore, we need to look closely at Tables A.5 and A.6 where we have a weak IV or no IV. We also see the pattern that standard deviations in Tables A.5 and A.6 are much smaller than alternative methods' standard deviations and they are quite the same as what we observed in Table A.18. In addition, we also note that the standard errors in Table A.18 use analytical variance instead of bootstrapping

variance which is not directly comparable to other methods' variances which are bootstrapping variances. However, the simulation result in Table A.13 implies that the analytical variance used in Table A.18 is quite reliable. For these reasons, we can conclude that the linear approximation may provide a poor approximation for the count and fractional response variable in a simultaneous model. It is also worth noting that standard deviations of nonlinear estimators are not increasing when we go from a strong to weak IV case. In this particular fractional probit model, too much of the identification appears to be off of the nonlinearity. Exclusion restriction seems not necessary when a nonlinear model is used for the first stage, instead of the linear model. In other words, nonlinearity is responsible for identification. However, it is widely considered preferable to have an identification strategy that is robust to using a linear first stage regression. This is really a matter of one's judgment and identification off the nonlinearity is still identification. We only need to worry about assumptions on functional form and the distribution of an error term.

1.5 Conclusion

I present the two-step QMLE and NLS methodology to estimate the fractional response model with a count endogenous explanatory variable. The unobserved heterogeneity is assumed to have an exponential gamma distribution, and the conditional mean of the fractional response model is estimated numerically. The two-step QMLE and NLS approaches are more efficient than the 2SLS and Tobit with IV estimates. They are more robust and less difficult to compute than the standard MLE method. This approach is applied to estimate the effect of fertility on the fraction of working hours for a female per week. Allowing the number of kids to be endogenous, using the data provided in Angrist & Evans (1998), I find that the marginal reduction of working hours per week is less as women have an additional kid. In addition, the effect of the number of children on the fraction of hours that a woman spends working per week is statistically significant and more significant than the estimates in all other linear and nonlinear models considered in this chapter.

Chapter 2

ESTIMATION OF A DYNAMIC TOBIT PANEL DATA WITH AN ENDOGENOUS VARIABLE AND AN APPLICATION TO FEMALE LABOR SUPPLY

2.1 Introduction

This chapter considers the estimation of a dynamic Tobit model with an endogenous regressor in the presence of unobserved heterogeneity in both stages and serial correlation in the first stage. Practical issue motivating this study is concerned with the dynamics of female annual labor supply where we have a corner solution outcome and it is affected by its previous state and another source of endogeneity. The estimation method proposed in this chapter is established based on a combination of practical methods proposed in the literature. To deal with the first source of endogeneity, estimation methods of dynamic nonlinear models with a lagged dependent variable have been proposed with fixed effects or random effects. The first method is case-specific, computationally complex and often leads to estimators that do not converge at the usual \sqrt{n} rate. In addition, partial effects for nonlinear model using this approach are not identified. Therefore, an appealing and robust method which solves for unobserved effects and the well-known initial condition problem has been proposed by Wooldridge (2005). To correct for the second source of endogeneity, we can use a control function approach, especially convenient and computationally easy proposed by Smith & Blundell (1986) for limited dependent variables (LDV). As state dependence in a dynamic nonlinear model can be overestimated without taking into account serial correlation, we also need to correct for serial correlation.

The contribution of this chapter is to provide a computationally attractive estimation method for a dynamic censored model with an endogenous regressor (besides the lagged dependent variable) and serially correlated error terms. This method is readily applied to Panel Study of Income Dynamics (PSID) data using the years 1980 to 1992. Based on the estimation result, I find the evidence of persistence in US white female labor working hours over the period 1980-1992. It suggests that the current labor supply of US women is affected by their past labor supply and their initial condition of labor supply. Both observed and unobserved individual heterogeneity, and serial correlation play an important role in the persistence of US female labor supply.

This chapter is organized as follows. The second section reviews: i) approaches to estimation of a dynamic Tobit panel data model; ii) a control function approach to govern the endogeneity problem and iii) methods to deal with serial correlation. It also discusses related issues on the dynamics of the US female annual labor supply. The third section develops a model for dynamic Tobit panel data with an endogenous regressor and the fourth section obtains average partial effects (APE) estimates. The fifth section discusses how to correct for serial correlation in the first stage. Empirical example follows in the next section. The last section is summarization and conclusion.

2.2 Literature Review

The approach and framework of this chapter are most closely related to the work proposed by Giles & Murtazashvili (2010). They allow continuous, endogenous contemporaneous regressors in a dynamic panel data model but their outcome of interest is a binary variable. Their estimator is applied to analyze the impact of migrant labor markets on reducing the probability of falling into poverty. Since the outcome variable in this chapter is continuous with a positive probability as well as has a pileup at zero, the framework for a dynamic binary response model in Giles & Murtazashvili (2010) has to be adjusted. A dynamic Tobit panel data model should be appropriate in this case.

There have recently been many studies on a dynamic Tobit panel data model which allows for unobserved heterogeneity and dynamic feedback. These two features of the dynamic panel data model, however, often create difficulties in estimation. The main difficulty is that with nonlinearity, it is not obvious how to "difference away" the individual specific effects and how to use instrumental variable type techniques.

Some developments have been made on estimating certain nonlinear dynamic models using the "fixed-effects" approach, for example, the censored regression models (Honore (1993); Hon-

ore & Hu (2001)), the sample selection models (Kyriazidou (2001)), the discrete choice models (Honore & Kyriazidou (2000)), and the models with multiplicative individual effects (Chamberlain (1992); Wooldridge (1997)). In particular, Honore (1993) proposed some solutions for estimating a censored regression panel data model with individual fixed effects and lagged censored dependent variables. Honore & Hu (2001) provided identification results for this approach under certain conditions. And Honore & Hu (2004) allowed a lagged dependent variable and a set of strictly exogenous variables. They constructed moment conditions for the panel data model with fixed effects and lagged (censored) dependent variable with a restrictive assumption of non-negative coefficient on the lagged dependent variable. In addition, their approach will not result in estimates for APEs. Even though semiparametric approaches do not make any assumptions on either unobserved effects or initial conditions but they are case-sensitive and often lead to estimators that do not converge at the usual \sqrt{n} rate (Arellano & Honore (2001)). For example, Honore & Kyriazidou (2000) assumed that transitory errors are iid over time (c_i is arbitrary dependent on X_{it}). If the regressors are continuous or have high dimension then the estimator will have a convergence rate slower than \sqrt{n} . The estimator will over-difference the data and understate the role of the initial value of dependent variable, y_{i0} . This causes downward biased coefficient on the lagged dependent variable in finite samples and this bias will not decrease as T increases. More importantly, partial effects on the conditional mean are not identified. The amount of state dependence therefore cannot be determined.

An alternative of estimation method in nonlinear dynamic models is to use the "randomeffects" approach. This approach is faced with a notably difficult issue of initial condition problem. Wooldridge (2005), Chay & Hyslop (1998), and Hsiao (1986) have an excellent summarization on how this problem is treated in the literature. There are three alternative assumptions on initial conditions. The first approach treats the initial condition as exogenous (Heckman (1978a,b, 1981b)). Initial conditions are independent of the individual effects and can be ignored when estimating the structural model. However, if either c_i or X_{it} is a determining factor in the initial sample conditions, then this approach will overstate the amount of state dependence in the process. Moreover, this is a very strong assumption and may not make sense. For example, ability is allowed to be uncorrelated with initial earnings. The second approach treats the initial condition as in equilibrium (Card & Sullivan (1988)). This restriction is unlikely to hold when observable covariates are time-varying and important determinants of the outcome. The initial condition is allowed to be random and the distribution of the initial condition given unobserved heterogeneity is specified. This model does not allow for additional covariates (Bhargava & Sargan (1983); Hsiao (1986)). The third way is to adopt a flexible reduced form specification: approximating initial sample observation (Heckman (1981b)). This approach is computationally difficult to obtain estimates of parameters and APEs.

The first approach is viewed as "pure" random effect approach where c_i is independent of z_i and y_{i0} . In addition, unobserved effect is independent of exogenous variables. One can obtain the density of (y_{i1}, \ldots, y_{iT}) given y_{i0} and \mathbf{z}_i by integrating out c_i . This method requires a strong assumption of independence between the initial condition and the unobserved effect. The fourth approach which is proposed by Wooldridge (2005) is the most unrestricted random effect model, which was named "correlated" random effect. Compared with the fixed effect model, it may provide substantial efficiency gain (Hausman (1978)) given the correctly specified distribution of c_i and y_{i0} . It recommends to obtain a joint distribution of (y_{i1}, \ldots, y_{iT}) conditional on y_{i0} and z_i ; rather than a distribution of (y_{i0}, \ldots, y_{iT}) conditional on \mathbf{z}_i as in Heckman's approach. However, we need to specify a density of c_i given y_{i0} and z_i (motivated by the original idea from Chamberlain (1980)). The relationship between c_i and \mathbf{z}_i makes this model named "correlated" random effects where we allow a linear relationship between c_i and z_i and y_{i0} . This approach requires fewer computational efforts than Heckman's technique and gives nice APEs. It also leads to several advantages, we can choose a flexible conditional distribution of the initial condition instead of approximation which results in computational difficulty. As a consequence, estimates are readily computed and partial effects can be easily determined.

The study of limited dependent variables models with an endogenous regressor (instead of lagged dependent variable) has a fairly long history. Most papers in the literature have assumed a reduced form for the endogenous variable. Examples of this include the papers by Nelson & Olson

(1978); Heckman (1978a); Amemiya (1979); Newey (1985, 1986); Blundell & Smith (1989); Vella (1993); Blundell & Powell (2004); Das (2002) for cross sections and Vella & Verbeek (1999); Labeaga (1999); Giles & Murtazashvili (2010) for panel data. In a linear model, such a reduced form (or the "first stage") can be thought of as a linear projection, and as such it is essentially always well-defined and consistently estimated by the OLS estimator. This is not the case in a nonlinear model where it is typically assumed that the first stage is a conditional expectation and that the error is independent of the instruments. Smith & Blundell (1986) considered a static Tobit model to analyze female labor supply in the UK in 1981 treating other household income as endogenous. The insight of their paper is to substitute a consistent estimator for the residual in the reduce-form equation into the structural model to control for the endogeneity. And this approach is named control function approach which produces a two-step estimator based on the conditional likelihood for the equation of interest. As this chapter studies a source of endogeneity not coming from the lagged dependent variable, we will employ the control function approach that allows for a correlation between unobserved effect and regressors, as well as between regressors and the structural error.

As in Baltagi & Li (1991) and Baltagi & Wu (1999), estimation of a panel data model with AR(1) disturbances is based on a feasible generalized least squares procedure. This method is simple to compute and provides natural estimates of the serial correlation and variance components parameters. The test for zero first-order serial correlation is also easily implemented. However, this estimation procedure works very well for a linear panel data model but has not been executed in a nonlinear panel data model. In order to deal with serial correlation in a dynamic nonlinear model, Lee (1999) has proposed the simulated maximum likelihood method. This method is robust in time-series context, however, it is quite computationally intensive. We will exploit the method similar to Baltagi & Li (1991) to handle our first stage serial correlation.

One of the possible applications for this model is to study the persistence of the US female labor supply taking into account endogenous features of observed covariates. The literature on labor supply has examined female labor supply in many studies. Women's labor supply is one of long-standing labor supply research (Heckman (1974); Heckman & Macurdy (1980)). Studies of women labor supply is growing rapidly due to the increasing availability of panel data and improved computational power and techniques. According to Heckman (1981a), state dependence may arise if working leads to accumulation of human capital – skills, know-how, work ethic, etc., and not working leads to depreciation of human capital. Women who prefer work to leisure, who are highly motivated and have high ability tend to stay in the work force for their entire working life and their high labor supply persistence is exhibited. Differences in "search costs" associated with different labor market states may also cause state dependence (Eckstein & Wolpin (1990); Hyslop (1999)). There might be fixed cost to enter the labor market, raising the cost for individuals who are not employed, relative to those already in the labor market.

Shaw (1994) studies the persistence of the US white female labor supply from 1967 to 1987 using a linear dynamic model with age stratification and she found persistence in their labor supply because as women entered the labor force, they tended to become continuous workers. She also found that the extent of persistence changed little over the 20 year period studied after controlling for individual circumstances which are influential for early and late life periods such as number of children, health status, age, and wages. However, she does not take into account the nature of the working hours as a limited dependent variable. And she does not examine whether the persistence comes from transitory shocks that might be serially correlated.

2.3 Model

I consider a panel data model with the latent variable as follows:

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + c_{1i} + u_{1it}, \quad t = 1, \dots, T,$$
(2.1)

$$y_{1it} = \max(0, y_{1it}^*), \tag{2.2}$$

where y_{1it} is observed and equal to zero with a positive probability while continuously distributed over strictly positive values, $y_{1i,t-1}$ is a lagged dependent variable and the dynamics are assumed first order, y_{2it} is an endogenous variable, \mathbf{x}_{it} is a 1 × K vector of time-varying explanatory variables which can contain a constant term, c_{i1} is a time-constant unobserved heterogeneity and u_{it1} is an idiosyncratic error. β is a $K \times 1$ vector of parameters, ρ and α are scalar parameters. *i* indexes a random draw from the cross section with sample size *N* and *t* denotes a particular time period within a number of fixed time periods *T*. For simplicity, we assume a balanced panel. In the followings, we have i = 1, ..., N, and t = 1, ..., T.

We assume that model (2.1) is correctly specified dynamically and the error term is serially uncorrelated:

$$u_{1it}|y_{1i,t-1},\ldots,y_{1i0},\mathbf{x}_i,c_{1i} \sim Normal(0,\sigma_{u_1}^2).$$
 (2.3)

If we allow the error term to be serially correlated, for example, allowing for an AR(1) process, we would want to include not only a lagged dependent variable but also lags of x as well. In this case, we include a single lag of y_1 , contemporaneous y_2 and possibly of x's.

In model (2.1), \mathbf{x}_{it} is assumed to be strictly exogenous and y_{2it} is allowed to be endogenous. Let $\mathbf{z}_{it} = (\mathbf{x}_{it}, \mathbf{z}_{1it})$ be a set of strictly exogenous variables, a $1 \times L$ vector of instrumental variables, where L > K and \mathbf{z}_{1it} is excluded from (2.1). Using the control function approach to model the endogeneity (see Smith & Blundell (1986); Rivers & Vuong (1988)), we can assume a linear reduced form for y_{2it} as follows:

$$y_{2it} = \mathbf{z}_{it}\gamma + c_{2i} + u_{2it}, \qquad (2.4)$$

where u_{2it} is an idiosyncratic serially uncorrelated error with $Var(u_{2it}) = \sigma_{u_2}^2$ and c_{2i} is an unobserved effect. Using Mundlak (1978)'s device, we allow $c_{2i} = \overline{z}_i \delta + a_{2i}$ and rewrite y_{2it} as:

$$y_{2it} = \mathbf{z}_{it}\gamma + \overline{\mathbf{z}}_i\delta + v_{2it}, \qquad (2.5)$$

where $v_{2it} = a_{2i} + u_{2it}$; $\overline{\mathbf{z}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{z}_{it}$ and $a_{2i} | z_i \sim Normal(0, \sigma_{a_2}^2)$. We can also add time dummies into this reduced form.

Now it boils down to the assumption that we need to make for the conditional distributions of u_{1it} and c_{1i} . We will discuss first about u_{1it} . As in the cross-sectional case discussed in Smith & Blundell (1986), we can allow a joint normality between u_{1it} and v_{2it} . However, v_{2it} is serially correlated because of the presence of the heterogeneity, a_{2i} , therefore, this will make the serial

correlation issue in the estimation with dynamics difficult to handle. As a result, we will start with the assumption of joint normality of u_{1it} and u_{2it} as suggested in Giles & Murtazashvili (2010) since u_{2it} can be naturally assumed to be serially uncorrelated. As we will see in the discussion below, this assumption is reasonable in our context of dynamics in the structural equation. We write:

$$u_{1it} = \theta_1 u_{2it} + e_{1it}, \tag{2.6}$$

where $\theta_1 = Cov(u_{1it}, u_{2it})/Var(u_{2it})$ and $e_{1it} \sim Normal(0, \sigma_{e1}^2)$.

 (u_{1it}, u_{2it}) is allowed to have a zero mean, bivariate normal distribution; \mathbf{z}_i is strictly exogenous in both equation (2.1) and (2.5) or in other words, u_{1it} and u_{2it} are independent of \mathbf{z}_i .

 e_{1it} is independent of \mathbf{z}_i and u_{2it} . We can assume that e_{1it} serially uncorrelated because u_{2it} is serially uncorrelated and independent of \mathbf{z}_i in addition to the fact that u_{1it} is free of serial correlation. Even if u_{2it} is serially correlated, we can correct for this serial correlation without any hardship. We will discuss about this issue in more details later.

Regarding the issue of the endogeneity of y_{2it} , let us rewrite equation (2.6):

$$u_{1it} = \theta_1 v_{2it} - \theta_1 a_{2i} + e_{1it}.$$
 (2.7)

We will see now the direct relation between u_{1it} and v_{2it} , through that we can account for endogeneity of y_{2it} in period t. In addition, u_{2it} is free of serial correlation, then from equation (2.6), e_{1it} is not correlated with $u_{2i,t-1}$ and $v_{2i,t-1}$, as a result. With the same idea for past values of v_{2it} , y_{2it} will become sequentially exogenous in the estimating equation. With equation (2.7), we now have to handle the heterogeneity issues, not only c_{1i} but also a_{2i} . Rewrite the structural equation under the assumption from equation (2.7), we have:

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + c_{1i} + \theta_1 v_{2it} - \theta_1 a_{2i} + e_{1it}, \quad t = 1, \dots, T,$$

or

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + s_i + \theta_1 v_{2it} + e_{1it}, \quad t = 1, \dots, T,$$
(2.8)

where $s_i = c_{1i} - \theta_1 a_{2i}$, which is a composite error.

Using Wooldridge-Chamberlain's device (2005, 1980), with the motivation of "correlated" random effects dynamic model proposed by Wooldridge (2005) to handle the initial condition problem, we can specify s_i as a linear function of y_{1i0} and \mathbf{z}_i in order to use the standard random effects Tobit software without approximating the density function of s_i . However, now our regressors in equation (2.8) extends to include v_{2it} (which is not in \mathbf{z}_i), therefore, we will include \mathbf{v}_{2i} into the linear function that describe the relationship of s_i and the initial condition as well as explanatory variables in all time periods.

$$s_i = \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + a_{1i}, \tag{2.9}$$

where

$$a_{1i}|(y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}) \sim Normal(0, \sigma_{a_1}^2).$$
 (2.10)

This is a reasonable assumption because unobserved effect (such as motivation, ambition) is correlated with the initial condition of the outcome of interest (working hours). In addition, as the model has a lagged dependent variable, $y_{1i,t-1}$ and c_{1i} has some source of correlation. In order to conserve the degree of freedom or to reduce the time of computation which will be important in some applied work with a substantial number of explanatory variables, we can assume that different time periods of explanatory variables have equal impacts on s_i and using Mundlak's device, we can restrict our assumption (2.9) to: $s_i = \theta_2 y_{1i0} + \overline{z}_i \theta_3 + \overline{v}_{2i} \theta_4 + a_{1i}$.

We can see that now $v_{1it} = \theta_2 y_{i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i} + e_{1it}$. Substitute that into equation (2.1), hence, we readily obtain:

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + \theta_2 y_{1i0} + \mathbf{z}_i\theta_3 + \mathbf{v}_{2i}\theta_4 + \theta_1 v_{2it} + a_{1i} + e_{1it}, \quad t = 1, \dots, T, \quad (2.11)$$

and in a shorter version, we have:

$$y_{1it} = \max(0, \mathbf{w}_{1it}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i} + e_{1it}), \quad t = 1, \dots, T,$$
(2.12)

where $\mathbf{w}_{1it} = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it})$ and $\lambda_1 = (\rho, \alpha, \beta')'$.

Based on the estimating equation (2.11), with the framework suggested in (Wooldridge, 2002, section 13.9) and Wooldridge (2005), we can write the density as follows:

$$f_t(y_{1t}|y_{1,t-1}, y_{2t}, y_{10}, \mathbf{z}, \mathbf{v}_2, a_1; \lambda) = f_{0t}f_{1t},$$
(2.13)

where

$$f_{0t} = \left\{1 - \Phi[(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{10} + \mathbf{z}\theta_3 + \mathbf{v}_2\theta_4 + \theta_1 v_{2t} + a_1)/\sigma_{e_1}]\right\}^{1|y_{1t}=0|}$$

and

$$f_{1t} = (1/\sigma_{e_1})\phi[(y_{1t} - \mathbf{w}_{1t}\lambda_1 - \theta_2 y_{10} - \mathbf{z}\theta_3 - \mathbf{v}_2\theta_4 - \theta_1 v_{2t} - a_1)/\sigma_{e_1}]^{1[y_{1t} > 0]}.$$

Thus the density of $(y_{1i1}, y_{1i2}, \dots, y_{1iT})$ given $(y_{1i0} = y_{10}, \mathbf{z}_i = \mathbf{z}, \mathbf{v}_{2i} = \mathbf{v}_2, a_{1i} = a_1)$ is:

$$\prod_{t=1}^{T} f_t(y_{1t}|y_{1,t-1}, y_{2t}, \mathbf{x}_t, y_{10}, \mathbf{z}, \mathbf{v}_2, a_1; \lambda),$$
(2.14)

and since we do not observe a_{1i} , in order to estimate λ , we need to integrate out a_1 from this density. Given $a_{1i}|(y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}) \sim Normal(0, \sigma_{a_1}^2)$, we can obtain the density of $(y_{1i1}, y_{1i2}, \dots, y_{1iT})$ given $(y_{1i0} = y_{10}, \mathbf{z}_i = \mathbf{z}, \mathbf{v}_{2i} = \mathbf{v}_2)$ as:

$$\int_{R} \left[\prod_{t=1}^{T} f_t(y_{1t}|y_{1,t-1}, y_{2t}, \mathbf{x}_t, y_{10}, \mathbf{z}, \mathbf{v}_2, a_1; \lambda) \right] (1/\sigma_{a_1}) \phi(a_1/\sigma_{a_1}) da_1,$$
(2.15)

which has exactly the same structure as in the standard random effects Tobit model, but the explanatory variables at time period *t* are:

$$\mathbf{w}_{it} = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it}, y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}, v_{2it}).$$
(2.16)

Now we can exploit the standard random effects Tobit software for estimation. We add y_{i0} , \mathbf{z}_i , $\hat{\mathbf{v}}_{2i}$ and \hat{v}_{2it} as additional explanatory variables in each time period and estimate λ , θ_3 , θ_4 and $\sigma_{e_1}^2$, where $\hat{\mathbf{v}}_{2i} = (\hat{\mathbf{v}}_{2i1}, \hat{\mathbf{v}}_{2i2}, \dots, \hat{\mathbf{v}}_{2iT})$.

Based on the above model development, the estimation procedure for "correlated random effect" dynamic Tobit model is proposed as follows.

Estimation Procedure:

- (i) Estimate the reduced form for y_{2it} using the pooled OLS of y_{2it} on \mathbf{z}_{it} , $\mathbf{\overline{z}}_{i}$, and time dummies. Obtain the residuals, $\hat{\mathbf{v}}_{2i}$ and \hat{v}_{2it} .
- (ii) Use the random effect Tobit of y_{1it} on $\widehat{\mathbf{w}}_{it}$ and get all the estimates of interest, $\widehat{\lambda}$, where $\widehat{\mathbf{w}}_{it} = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it}, y_{1i0}, \mathbf{z}_i, \widehat{\mathbf{v}}_{2i}, \widehat{\mathbf{v}}_{2it}).$

2.4 Average Partial Effects

In order to compare the magnitude of the estimate obtained in a nonlinear model from the previous section with a linear estimate, we need to obtain the marginal effect or the average partial effect (APE) of the explanatory variable of interest. Following Wooldridge (2002, 2005), the APEs are computed as the derivatives or differences of:

$$E[m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i}, \sigma_{e_1}^2)], \quad t = 1, \dots, T,$$
(2.17)

where $m(\mathbf{g}, \sigma_{e_1}^2) = \Phi[\mathbf{g}/\sigma_{e_1}]\mathbf{g} + \sigma_{e_1}\phi[\mathbf{g}/\sigma_{e_1}]$ under the notation that

$$\mathbf{g} = \mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i},$$

and in the argument of the expectation operator, variables with a subscript i are random and all others are fixed.

Using iterated expectation, expression (2.17) can be rewritten as:

$$E\{E[m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i\theta_3 + \mathbf{v}_{2i}\theta_4 + \theta_1 v_{2it} + a_{1i}, \sigma_{e_1}^2)|y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}]\},$$
(2.18)

where \mathbf{w}_{1t} are fixed values here and the conditional expectation is with respect to the distribution of $(y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}, a_{1i})$. Since a_{1i} and $(y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i})$ are independent, and $a_{1i} \sim Normal(0, \sigma_{a_1}^2)$, the conditional expectation in equation (2.18) is obtained by integrating

$$m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i}, \sigma_{e_1}^2),$$

over a_{1i} with respect to the Normal $(0, \sigma_{a_1}^2)$ distribution.

Since

$$m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i\theta_3 + \mathbf{v}_{2i}\theta_4 + \theta_1 v_{2it} + a_{1i}, \sigma_{e_1}^2)$$

is obtained by integrating

$$\max(0, \mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i} + e_{1it})$$

with respect to e_{1it} over the Normal $(0, \sigma_{e_1}^2)$ distribution, the conditional expectation in equation (2.18) is:

$$m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it}, \sigma_{a_1}^2 + \sigma_{e_1}^2).$$
(2.19)

For a given value of \mathbf{w}_{1t} (\mathbf{w}_1^0), a consistent estimator for expression (2.19) can be obtained by replacing unknown parameters by consistent estimators:

$$N^{-1}\sum_{i=1}^{N} m(\mathbf{w}_{1}^{0}\hat{\lambda}_{1} + \hat{\theta}_{2}y_{1i0} + \mathbf{z}_{i}\hat{\theta}_{3} + \hat{\mathbf{v}}_{2i}\hat{\theta}_{4} + \hat{\theta}_{1}\hat{v}_{2it}, \hat{\sigma}_{a_{1}}^{2} + \hat{\sigma}_{e_{1}}^{2}), \qquad (2.20)$$

where \hat{v}_{2it} are the first stage pooled OLS residuals from y_{2it} on \mathbf{z}_{it} , $\overline{\mathbf{z}}_i$ and time dummies, and $\hat{\mathbf{v}}_{2i} = (\hat{v}_{2i1}, \hat{v}_{2i2}, \dots, \hat{v}_{2iT})$.

The APEs are obtained by taking derivatives or differences of expression (2.19) (in which \mathbf{w}_1^0 is replaced with \mathbf{w}_{1t}) with respect to \mathbf{w}_{1t} and the estimator of these APEs will be obtained based on those derivatives and differences and estimated parameters.

For example, \widehat{APE} of $y_{1,t-1}$ is:

$$\hat{\rho}\left\{(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\Phi[(\mathbf{w}_{1t}\hat{\lambda}_{1}+\hat{\theta}_{2}y_{1i0}+\mathbf{z}_{i}\hat{\theta}_{3}+\hat{\mathbf{v}}_{2i}\hat{\theta}_{4}+\hat{\theta}_{1}\hat{v}_{2it})/(\hat{\sigma}_{a_{1}}^{2}+\hat{\sigma}_{e_{1}}^{2})]\right\},$$
(2.21)

and \widehat{APE} of y_{2t} is:

$$\widehat{\alpha}\left\{ (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \Phi[(\mathbf{w}_{1t} \hat{\lambda}_{1} + \hat{\theta}_{2} y_{1i0} + \mathbf{z}_{i} \hat{\theta}_{3} + \hat{\mathbf{v}}_{2i} \hat{\theta}_{4} + \hat{\theta}_{1} \hat{v}_{2it}) / (\hat{\sigma}_{a_{1}}^{2} + \hat{\sigma}_{e_{1}}^{2})] \right\}.$$
(2.22)

2.5 Serial Correlation Correction

As discussed in the previous part, the essential assumption that we made in equation (2.6) requires u_{2it} free of serial correlation. If u_{2it} is serially correlated, then we must correct for the serial

correlation in e_{1it} , otherwise our estimator will not be consistent. For simplicity, assume that u_{2it} follows an AR(1) process, similar to the discussion in Giles & Murtazashvili (2010):

$$u_{2it} = \eta u_{2i,t-1} + e_{2it}, \quad t = 1, \dots, T,$$
(2.23)

and e_{2it} is a white noise error with $Var(e_{2it}) = \sigma_{e_2}^2$.

We have:

$$y_{2it} = \mathbf{w}_{2it} \gamma_2 + a_{2i} + u_{2it}, \tag{2.24}$$

where $\mathbf{w}_{2it} = (\mathbf{z}_{it}, \overline{\mathbf{z}}_i)$ and $\gamma_2 = (\gamma', \delta')'$ or we can write:

$$y_{2i} = \mathbf{w}_{2i}\gamma_2 + v_{2i}.\tag{2.25}$$

Since u_{2it} has serial correlation, e_{1it} is serially correlated as we can see below:

$$\begin{split} \eta_{e} &= Cov(e_{1it}, e_{1i,t-1}) = Cov(u_{1it} - \theta_{1}u_{2it}, u_{1i,t-1} - \theta_{1}u_{2i,t-1}), \\ \eta_{e} &= Cov(u_{1it} - \theta_{1}\eta_{u_{2i,t-1}} - \theta_{1}e_{2it}, u_{1i,t-1} - \theta_{1}u_{2i,t-1}), \\ \eta_{e} &= \eta \theta_{1}^{2} Var(u_{2i,t-1}), \\ \eta_{e} &\neq 0 \text{ unless } \eta = 0 \text{ or } \theta_{1} = 0. \end{split}$$

To remove serial correlation in e_{1it} , our strategy is to use a transformation procedure and obtain the first-stage residual free of serial correlation. Define the variance-covariance matrix of \mathbf{v}_{2i} as:

$$\Gamma = E(\mathbf{v}_{2i}\mathbf{v}_{2i}') = \sigma_{a_2}^2 \mathbf{j}_T \mathbf{j}_T' + \sigma_{u_2}^2 \Psi(\eta), \qquad (2.26)$$

where Γ is a $T \times T$ positive definite matrix when $-1 < \eta < 1$ and I assume that in what follows. This matrix is necessarily the same for all *i* because of the random sampling assumption in the cross section. **j**_{*T*} is a $T \times 1$ vector of ones, and $\Psi(\eta)$ is defined as below:

$$\Psi(\eta) = \begin{bmatrix} 1 & \eta & \eta^2 & \dots & \eta^{T-3} & \eta^{T-2} & \eta^{T-1} \\ \eta & 1 & \eta & \dots & \eta^{T-4} & \eta^{T-3} & \eta^{T-2} \\ \eta^2 & \eta & 1 & \dots & \eta^{T-5} & \eta^{T-4} & \eta^{T-3} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \eta^{T-2} & \eta^{T-3} & \eta^{T-4} & \dots & \eta & 1 & \eta \\ \eta^{T-1} & \eta^{T-2} & \eta^{T-3} & \dots & \eta^2 & \eta & 1 \end{bmatrix}.$$
(2.27)

We also note that $\sigma_{u_2}^2 = \sigma_{e_2}^2/(1-\eta^2)$. After obtaining consistent estimates of η , $\sigma_{a_2}^2$, $\sigma_{u_2}^2$ (and $\sigma_{e_2}^2$), we can transform v_{2it} into v_{2it}^* which is free of serial correlation. With this new serially uncorrelated error (v_{2it}^*) , we can transform u_{2it} to a new serially uncorrelated u_{2it}^* using this equation: $u_{2it}^* = v_{2it}^* - a_{2i}$. This will guarantee that our new e_{1it} (e_{1it}^*) is free of serial correlation as a result of: $e_{1it}^* = u_{1it} - \theta_1 u_{2it}^*$. e_{1it}^* is now serially uncorrelated, independent of \mathbf{z}_i and u_{2it}^* , and has a normal distribution: $Normal(0, \sigma_{e_1}^{*2})$.

We will briefly describe the transformation procedure as follows:

Using the fact that $\mathbf{j}_T' \mathbf{j}_T = T$, Γ is rewritten as:

$$\Gamma = T \sigma_{a_2}^2 \mathbf{j}_T \left(\mathbf{j}_T' \mathbf{j}_T \right)^{-1} \mathbf{j}_T' + \sigma_{u_2}^2 \Psi(\eta) = T \sigma_{a_2}^2 \mathbf{P}_T + \sigma_{u_2}^2 \Psi(\eta), \qquad (2.28)$$

where $\mathbf{P}_T \equiv \mathbf{I}_T - \mathbf{Q}_T$; $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{j}_T (\mathbf{j}_T' \mathbf{j}_T)^{-1} \mathbf{j}_T'$. Define $\tau_1 = \sigma_{u_2}^2 \Psi(\eta) / [T \sigma_{a_2}^2 + \sigma_{u_2}^2 \Psi(\eta)]$, we can write:

$$\Gamma = \left[T\sigma_{a_2}^2 + \sigma_{u_2}^2\Psi(\eta)\right](\mathbf{P}_T + \tau_1\mathbf{Q}_T).$$
(2.29)

After some algebra, we can show that: $(\mathbf{P}_T + \tau_1 \mathbf{Q}_T)^{-1/2} = (1 - \tau)^{-1} [\mathbf{I}_T - \tau \mathbf{P}_T]$ where $\tau = 1 - \sqrt{\tau_1}$.

Hence,

$$\Gamma^{-1/2} = \left[T \sigma_{a_2}^2 + \sigma_{u_2}^2 \Psi(\eta) \right]^{-1/2} (1 - \tau)^{-1} [\mathbf{I}_T - \tau \mathbf{P}_T] = \left[\sigma_{u_2}^2 \Psi(\eta) \right]^{-1/2} [\mathbf{I}_T - \tau \mathbf{P}_T], \quad (2.30)$$

where $\tau = 1 - \left\{ \sigma_{u_2}^2 \Psi(\eta) / [T \sigma_{a_2}^2 + \sigma_{u_2}^2 \Psi(\eta)] \right\}^{1/2}.$

Define $\mathbf{C}_T \equiv [\Psi(\eta)]^{-1/2} [\mathbf{I}_T - \tau \mathbf{P}_T]$ and transform equation (2.25) into:

$$\widetilde{\mathbf{y}}_{2i} = \widetilde{\mathbf{w}}_{2i} \gamma_2 + \widetilde{\nu}_{2i}, \tag{2.31}$$

by multiplying C_T to both sides of equation (2.25).

Now the variance matrix of \tilde{v}_{2i} is:

$$E(\tilde{v}_{2i}\tilde{v}'_{2i}) = \mathbf{C}_T \Gamma \mathbf{C}_T = \sigma_{u_2}^2 \mathbf{I}_T.$$
(2.32)

Therefore we have transformed \mathbf{v}_{2i} into $\mathbf{v}_{2i}^* (= \tilde{v}_{2i})$ which is serially uncorrelated and homoskedastic by using:

$$\widetilde{v}_{2i} = \mathbf{C}_T \mathbf{v}_{2i}.\tag{2.33}$$

The estimator of C_T is:

$$\widehat{\mathbf{C}}_T = \widehat{\sigma}_{u_2} \widehat{\Gamma}^{-1/2}. \tag{2.34}$$

We can see that, in the special case when $\eta = 0$ (no serial correlation), $\Psi(\eta) = I_T$ and $C_T = [I_T - \tau P_T]$.

Now we can adjust equation (2.9) under the adjusted assumption that:

$$s_i^* = \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i}^* \theta_4 + a_{1i}^*, \qquad (2.35)$$

where $a_{1i}^*|(y_{i0}, \mathbf{z}_i, \mathbf{v}_{2i}^*) \sim Normal(0, \sigma_{a_1}^{*2})$ and obtain:

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + \theta_2 y_{1i0} + \mathbf{z}_i\theta_3 + \mathbf{v}_{2i}^*\theta_4 + \theta_1 v_{2it}^* + a_{1i}^* + e_{1it}^*, \quad t = 1, \dots, T, \quad (2.36)$$

where $\mathbf{v}_{2i}^* = (\mathbf{v}_{2i1}^*, \mathbf{v}_{2i2}^*, \dots, \mathbf{v}_{2iT}^*)$ and we will estimate all parameters in the second stage using standard random effects Tobit software, based on the density of $(y_{1i1}, y_{1i2}, \dots, y_{1iT})$ given $(y_{1i0} = y_{10}, \mathbf{z}_i = \mathbf{z}, \mathbf{v}_{2i}^* = \mathbf{v}_2^*)$ as:

$$\int_{R} \left[\prod_{t=1}^{T} f_t(y_{1t}|y_{1,t-1}, y_{2t}, \mathbf{x}_t, y_{10}, \mathbf{z}, \mathbf{v}_2^*, a_1^*; \lambda^*) \right] (1/\sigma_{a_1}^*) \phi(a_1^*/\sigma_{a_1}^*) da_1^*,$$
(2.37)

which has exactly the same structure as in the standard random effects Tobit model, but the explanatory variables at time period *t* are:

$$\mathbf{w}_{it}^* = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it}, y_{1i0}, \mathbf{z}_i, \mathbf{v}_{2i}^*, v_{2it}^*).$$
(2.38)

Now we can propose an estimation procedure for "correlated random effect" dynamic Tobit model with first-stage residual serial correlation correction.

2.5.1 Estimation Procedure

- (i) Run the random effect linear regression with an AR(1) disturbance of y_{2it} on \mathbf{w}_{2it} (with time dummies) and obtain the residuals \hat{v}_{2it} and $\hat{\mathbf{v}}_{2i}$. Obtain $\widehat{\mathbf{C}}_T$ and transform \hat{v}_{2it} and $\hat{\mathbf{v}}_{2i}$ into \hat{v}_{2it}^* and $\hat{\mathbf{v}}_{2i}^*$ based on the above transformation procedure.
- (ii) Use the random effect Tobit of y_{1it} on $\widehat{\mathbf{w}}_{it}^*$ and get all the estimates of interest, $\widehat{\lambda}^*$, where $\widehat{\mathbf{w}}_{it}^* = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it}, y_{1i0}, \mathbf{z}_i, \widehat{\mathbf{v}}_{2i}^*, \widehat{\mathbf{v}}_{2it}^*).$

2.5.2 Average Partial Effects

As the errors in the first stage are serially correlated, we also need to adjust the estimates of APEs. Instead of equation (2.18), we start with:

$$E[m(\mathbf{w}_{1t}\lambda_1 + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i}^* \theta_4 + \theta_1 v_{2it}^* + a_{1i}^*, \sigma_{e_1}^{2*})], \quad t = 1, \dots, T,$$
(2.39)

and following the same discussion as the case with no serial correlation, we can obtain APEs with respect to $y_{1,t-1}$, y_{2t} , and \mathbf{x}_t by taking derivatives or differences of:

$$N^{-1}\sum_{i=1}^{N} m(\mathbf{w}_{1i}\hat{\lambda}_{1} + \hat{\theta}_{2}y_{1i0} + \mathbf{z}_{i}\hat{\theta}_{3} + \hat{\mathbf{v}}_{2i}^{*}\hat{\theta}_{4} + \hat{\theta}_{1}\hat{v}_{2it}^{*}, \hat{\sigma}_{a_{1}}^{2*} + \hat{\sigma}_{e_{1}}^{2*}).$$
(2.40)

If the null hypothesis of no endogeneity and no serial correlation in the first stage is rejected, the standard errors in the second stage should be adjusted for the first stage estimation by using delta method or bootstrapping. In addition, we also need to obtain asymptotic standard errors for the APEs. Appendix E shows how to obtain adjusted standard errors in the second stage and asymptotic standard errors for the APEs using delta method.

2.5.3 Comparison

We will compare the methods proposed in the previous section with the traditional linear model and the model without serial correlation correction.

1. Linear Dynamic Model with an endogenous explanatory variable

We estimate model (2.1) using a generalized method of moments (GMM) system approach (Arellano & Bover (1995)) using both level and differenced instruments.

2. Correlated Random Effect model (without serial correlation correction)

We estimate model (2.1) with a correlated random effect model:

$$y_{1it}^* = \rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + \theta_2 y_{1i0} + \mathbf{z}_i \theta_3 + \mathbf{v}_{2i} \theta_4 + \theta_1 v_{2it} + a_{1i} + e_{1it}, \quad t = 1, \dots, T.$$

Estimation Procedure:

- (i) Estimate the reduced form for y_{2it} using the pooled OLS of y_{2it} on \mathbf{z}_{it} , $\mathbf{\overline{z}}_i$ and time dummies. Obtain the residuals, $\hat{\mathbf{v}}_{2i}$ and \hat{v}_{2it} .
- (ii) Use the random effect Tobit of y_{1it} on $\widehat{\mathbf{w}}_{it}$ and get all the estimates of interest, $\widehat{\lambda}$, where $\widehat{\mathbf{w}}_{it} = (y_{1i,t-1}, y_{2it}, \mathbf{x}_{it}, y_{1i0}, \mathbf{z}_i, \widehat{\mathbf{v}}_{2i}, \widehat{\mathbf{v}}_{2it})$ using the notation introduced in the previous section.

Using Mundlak's simpler version of Chamberlain's device (1980), in those estimating equations above, we can use $\overline{\mathbf{v}}_{2i}$ instead of \mathbf{v}_{2i} since the Mundlak's model can conserve on degrees of freedom, which is important especially when *T* is large in a dynamic model.

2.6 Empirical Example

The estimation procedure described above can be used in many applications. Here we apply to analyze the US female labor supply. In a panel data study, working hours exhibit a dynamic behavior and the persistence may be contaminated by heterogeneity, endogeneity and serial correlation. According to Heckman and MaCurdy's labor supply model, the censored model should be appropriate. The challenge is to invent a new econometric device to estimate a dynamic censored model with an endogenous variable besides the lagged dependent variable. And this new device has been developed in the previous section.

Endogeneity of experience is a potential problem because there are two sources of endogeneity here. First, experience is correlated with ability. Second, experience is constructed based on working hours and exogenous shock to working hours in the past (through wages) is correlated with the number of years of experience we observe today. Therefore, experience is not viewed as strictly exogenous after conditioning on unobserved heterogeneity.

Having controlled for unobserved effects does not follow that we have unbiased estimates of state dependence, for two reasons. First, women with high average lifetime hours of work, and thus high \mathbf{x}_{it} , may have become permanent workers because their early experience in the market demonstrated to them the need for continuous hours of work to build and maintain their human capital investment. The result is that "human capital acquired through work experience raises

the future probability of participation" (see Heckman & Willis (1974), initial definition of state dependence). In this case, the estimated state dependence parameter is biased downward towards 0, because state dependence operates entirely through a high lifetime c_i . State dependence is the coefficient on lagged hours: a positive coefficient on lagged hours implies that past hours have a positive impact on future hours. If c_i is omitted from the regression, the coefficient on lagged hours will be biased upward by the omitted variable bias, and therefore the importance of state dependence will be over-estimated.

The second problem is that state dependence cannot be separated from serially correlated errors. In other words, if shocks to hours are correlated over time, they will be picked up by the lagged hours variable, and state dependence will be biased. Hyslop (1999) found that transitory errors negatively correlated over time, suggesting failing to control for serially correlated transitory errors would lead to underestimation of state dependence.

2.6.1 Data

One application of the model introduced in the previous part is to study the dynamics of female labor supply. We can use the data from the Panel Study of Income Dynamics (PSID) for the years 1980-1992. In this study, we only focus on 864 white female who were either heads of households or spouses and their age is from 18 to 65. Women who are self-employed, in army and agricultural workers are excluded. Observations with inconsistent or missing data are dropped. More specifically, if one of the following happened in at least one year between 1976-1992, then the person will be dropped: self-reported age exceeded the age constructed using information on the year of birth by more than two years or self-reported age was smaller than constructed age by more than one year; the person was less than 18 or more than 65 years old; the person had missing experience; the person's age exceeded her/his experience by less than six years; spouse's weeks of unemployment was missing; the person reported positive work hours and zero earnings; the change in years of schooling between 1976-1985 was negative and exceeded one year in absolute value. In cases when the reported decrease in years of schooling was on year, the minimum of the

two reported values was assigned in all periods. The final sample consists of 11,232 observations.

The dependent variable in the structural equation (2.1), y_{1it} , is female annual working hours. The vector of explanatory variables includes the lagged dependent variable ($y_{1i,t-1}$), an endogenous variable, experience, (y_{2it}), and a set of exogenous variables (\mathbf{x}_{it}): education (measured in years of schooling), number of small children ages in 3 categories: 0-2, 3-5, and 6-17, marital status, husband's employment status, and non-wife income. Experience is constructed by taking the information about prior experience from 1976 survey year or from the year when the individual entered the sample for the first time, and then updating this information annually. In each year, experience was increased by one if the annual work hours were 2000 or more, and it was increased by the number of hours worked divided by 2000 if the annual work hours were less than 2000. Education is considered to be strictly exogenous conditional on the unobserved effect while experience is considered endogenous.

The set of instruments, z_{it} , contains years of schooling, age and its square, an indicator of marital status, number of children with three categories of ages in the family, husband's employment status, and non-wife income; their time averages and time dummies. Table B.1 reports the summary statistics for all variables used in the analysis.

Figure B.1 shows the distribution of women's working hours during the period 1980-1992. Around 27 percent of women did not work at the time of the survey. On average, women work for 1124 hours per year, which is about 21 hours per week (including women who do not work). The next largest group consists of women who work for 2000 hours per year, which is equivalent to 40 hours per week, accounting for 12 percent. The pattern with some pile up at zero hour and 2000 hours suggests that hours of work are sensitive to changes in the structure of both observed and unobserved individual heterogeneity. Figures B.2-B.5 illustrate the relationship between women's working hours and her experience, and her children number with 3 groups of ages, respectively. All of these relationships appear to fit our prior expectations.

In our sample, there are 2,978 women worked zero hours, opposed to 8,053 women worked for wage during the year with positive hours, ranging from 2 to 5,168. Hence, annual hours is a

reasonable candidate for a Tobit model.

2.6.2 Estimation and Result

We are interested in estimating the dynamic Tobit model of working hours for a woman *i* at time *t*:

$$Hours_{it} = \max(0, \rho Hours_{i,t-1} + \alpha Experience_{it} + \mathbf{x}_{it}\beta + c_{1i} + u_{1it}), \quad (2.41)$$

where $Hours_{it}$ is annual working hours for a woman *i* at time *t*, which are determined by her annual working hours in the previous period, $Hours_{i,t-1}$, her experience, $Experience_{it}$, and a vector of her characteristics including age, education, number of children, marital status and her husband's characteristics. The lagged dependent variable is included to capture the dynamic feature of working hours, in the sense that current working hours may also depend on past working hours, all others held constant. This dependence is due to things such as the accumulation of skills derived from past work. From this model, we are interested in estimating the coefficients on $Hours_{i,t-1}$ and $Experience_{it}$. The coefficient on $Hours_{i,t-1}$ will shed light on the US female labor supply persistence over the period 1980-1992.

As women's experience is considered endogenous in this model, we will instrument the endogenous regressor with her age and its square because there is a positive significant correlation between experience and age and age is strictly exogenous in the structural equation. The firststage regression estimates and their statistics are reported in Table B.2. The instruments are jointly significant on experience with the F-statistics are 196.26.

We first test for the endogeneity of fertility using the Hausman (1978) test. Because experience and working hours are simultaneously determined, the exogeneity assumption of experience has to be tested. A test for endogeneity of y_{2it} can be obtained by adding the first-stage residuals to the second stage estimation and obtain the t-statistic on \hat{v}_{2it}^* . Table B.3 shows the significance of \hat{v}_{2it} and \hat{v}_{2it}^* suggests that the hypothesis of an exogenous experience is rejected.

Table B.4 reports estimation results (of average partial effects) using the correlated random effect approach with and without serial correlation correction. The estimation result for the dynamic linear model using GMM method is also shown for comparison. Since the result in Table B.4 is consistent with the result in Table B.3, we are going to discuss more about the results in Table B.3. In all models (columns (1)-(3)), the coefficients for lagged working hours are significant and positive, suggesting positive state dependence of labor supply for women. The positive sign of the lagged working hours shows that women are likely to continue to be workers if they are already workers or continue to be unemployed if she does not work. The decline in the value of the coefficient on lagged working hours from model (1) to model (3) explains the upward bias of state dependence in women's working hours without taking into account the censored and unobserved heterogeneity issues as well as the serial correlation of unobserved factors. Unobserved heterogeneity which correlates with women characteristics contributes the largest to this upward bias, next is the ignorance of zero working hours issues and last is the serial correlation of unobserved factors.

In these models, from column (1) to column (3), in general, experience has positive influence on working hours. The magnitude is larger when we controlled for serial correlation. It shows that if women work continuously and accumulate a substantial amount of experience, the more experience they have, the more hours they work.

Compare to columns (1) and (2), we control for an extra source of serial correlation (the transitory shock) in experience besides unobserved heterogeneity (the permanent shock). The coefficient on experience is quite larger and its standard error is smaller. The intuition is as follows. Consider a positive (transitory) shock to experience. With a high degree of positive serial correlation and a rise of experience in the first period, experience will continue to rise in the next period and become very large over a long time period. This explains a higher coefficient on experience compared to those on (1) and (2). After correcting for the serial correlation in (3), even though in the first stage, CRESC estimates are more efficient than CRE estimates and we can see that the standard error on experience is smaller than those on (1) and (2).

Other explanatory variables might be affected using CRESC (for example, number of children) because when a lagged dependent variable entered into the equation, which is the proxy of the

dependent variable, in the presence of the serial correlation of the endogenous variable, the higher effect of experience may pick up some of the effects of unmeasured variables as well as observed covariates. As a result, the coefficients on the lagged dependent variable and children (as well as mother's education) are reduced and the significance of children might change. Even though a linear model does not require any serial correlation assumption of experience, the coefficient of children is more appealing in a nonlinear model where we use CRESC. We also note that children is allowed to be correlated with heterogeneity but not with the shocks to labor supply so this assumption is not conflicting with the endogeneity assumption in Chapter 1 where we deal with the cross section and allow correlation between children and heterogeneous preference. In this chapter, we treat children exogenous with respect to shocks rather than with respect to heterogeneity.

It is also indicated from the coefficients on small children from 0-2 and 3-5 that small kids have statistically significant negative effects on mothers' working hours. There is an evidence from the result that children aged 6 to 17 do not affect negatively to women working hours and the statistics are not significant in models (2) and (3).

The initial value of working hours illustrates the correlation between the unobserved effect and the initial condition. The coefficient on the initial value of working hours is statistically significant in both models (2) and (3). It suggests a strong state dependence of labor supply for women for a long period.

2.7 Conclusion

In this chapter, an attractive and easy-to-compute method for estimating dynamic Tobit panel data models with endogenous regressors (besides the lagged dependent variable) is proposed. This approach requires fewer computational efforts than Heckman's technique and gives nice APEs. It also leads to several advantages, for example, we can choose a flexible conditional distribution of the initial condition instead of approximation which results in computational difficulty. As a consequence, estimates are readily computed and partial effects can be easily determined. In addition, the control function approach is used to control for the endogeneity which is not coming

from the lagged dependent variable. This approach allows for correlation between unobserved effect and regressors, as well as between regressors and the structural error. To handle the presence of heterogeneity that causes serial correlation, the correction procedure is added and the serially uncorrelated residual in the first stage is obtained.

This proposed method discussed in this chapter provides useful tool for applied economic research. The method can be applied to various economic applications, such as estimation of labor supply models, housing expenditure models, or children's educational expenditure models, etc. The proposed estimation procedure is readily applied to Panel Study of Income Dynamics data from 1980 to 1992. Based on the estimation result, I find a strong evidence of persistence in the US white female labor working hours after controlling for censoring, endogeneity and serial correlation issues. I also find that the initial condition of female labor supply is statistically significant and has positive impact on women working history. It suggests that the current labor supply of US women is affected by their past labor supply and their initial condition of labor supply.

Chapter 3

AN EXPONENTIAL TYPE II TOBIT PANEL DATA MODEL WITH BINARY ENDOGENOUS REGRESSOR - APPLICATION TO ESTIMATING THE EFFECT OF FERTILITY ON MOTHERS' LABOR FORCE PARTICIPATION AND LABOR SUPPLY

3.1 Introduction

There has been a growing interest in the estimation of nonlinear panel data models with discrete endogenous variables. Most of the studies focus on binary response or count models with an endogenous dummy variable. However, there has not been any method suggested in a panel data model with a corner solution response. Moreover, there is a correlation between the probability of a positive outcome and itself. Heterogeneity is also present in the model. Therefore, the goal of this chapter is to develop a panel data estimation method for a model with a corner solution response and a binary endogenous variable in the presence of heterogeneity and the mentioned correlation.

Many approaches have been proposed to handle switching endogeneity in models with limited dependent variables. In a limited dependent variable panel data model, the main difficulty lies with the nonlinear functional form and we cannot difference away the unobserved effect. Fullinformation maximum likelihood can be used but this approach is intensively computational which makes it unattractive. Semiparametric or nonparametric estimators are based on distributional weaker assumptions; nevertheless, these estimators give scaled index coefficients and not average partial effects. The simplest approach is 2SLS, however, this method ignores nonlinearity in both the first and second stage. It might provide a good approximation but the two assumptions that a binary endogenous variable is expressed as a linear function and a binary or censored dependent variable is a linear function of a binary endogenous variable are unrealistic. Especially, this approach ignores the distribution of a censored variable where there is a massive pile of zeros. Econometricians came up with the control function approach to handle endogeneity so that nonlinearity is present in the second stage but linearity is still endured in the first stage. In this chapter, I propose a simple two-step estimator that keeps the nonlinearity assumption in both the first and second stage, and this method is more computationally attractive than the full-information maximum likelihood approach.

The model and estimation can be used in various economic applications. For example, we can apply it to study the effects of union status on labor market outcomes, the effects of childbearing on women's labor supply and many studies on health economics, business or epidemiology where binary endogeneity and the corner solution response occur. There are enormous studies on the effect of fertility on women's labor force participation (LFP) and labor supply. It is important to understand how the childbearing decision affects female participation in the labor force and how much she will work in a system of related equations. In this chapter, I will consider the fertility decision an endogenous dummy variable that influences both women's LFP and hours of work. The labor supply equation is the amount equation with a corner solution response while the LFP equation is the so-called participation equation. Using this system of equations, we can correct for both corner solution and endogenous problems in the study of women's labor supply.

The contribution of this chapter is to propose a simple two-step estimator which is robust and can be easily implemented for a Tobit panel data model in the presence of discrete endogeneity and heterogeneity. The main estimation strategy is to add correction terms so that the endogeneity and corner solution bias will be removed. This approach allows a joint distribution of the endogenous dummy regressor and the unobserved factors that affect both the amount and participation equations. I propose a two-step estimation method in which the first stage exploits a bivariate probit model for the relationship between the dummy endogenous variable and the participation decision. For the amount equation, by using an Exponential Type II Tobit (ET2T) model (see more of Type II Tobit models in (Wooldridge, 2010, chapter 17)), we can ensure that predicted value of log(hours) is positive, and there is a correlation between unobserved effects in both the amount and participation equation equation. In addition, exclusion restriction is used in order to identify the parameters in the structural equation. In other words, we allow some variables in the participation equation which

are not determinants in the amount equations. Explanatory variables are permitted to be correlated with the heterogeneity. Finally, on the empirical side, it also contributes to the study on the effect of having a newborn on women's LFP and labor supply, taking into account their unique culture and characteristics, using Vietnamese Household data in recent years.

This chapter is organized as follows: The second section reviews approaches to the estimation of a model with a binary endogenous explanatory variable and a limited dependent variable. It also discusses the literature on the effect of fertility on female labor supply and female labor participation. The third section develops a model for Tobit panel data with a dummy endogenous regressor in the presence of correlated participation and heterogeneity. An estimation procedure is proposed and average treatment effects are obtained. The next section gives an overview of data and estimation results for an empirical example. The last section is summarization and conclusion.

3.2 Literature Review

There have been many studies on limited dependent variable models with a dummy endogenous variable. These models were first pioneered by Heckman (1978a) using joint normal distributional assumptions and maximum likelihood (ML) method. Many other works use the conditional ML framework such as Amemiya (1978, 1979); Newey (1986, 1987); Blundell & Smith (1989) but with different procedures: generalized least squares (GLS) estimators, minimum Chi-squared estimators or two-step estimators. The disadvantage of this canonical method is that it is hard to implement and very computationally expensive.

In a panel data framework, most papers assume a reduced form for an endogenous variable or use a control function approach with generalized residuals (Vella & Verbeek (1999); Labeaga (1999)). Many studies also use this approach for cross-sectional cases (Vella (1993); Smith & Blundell (1986); Rivers & Vuong (1988)). Even though this approach produces consistent estimators, it would be unrealistic to assume a linear function for a dummy variable.

In order to avoid distributional assumptions in traditional ML framework as in Heckman (1978a), some studies have proposed nonparametric or semiparametric estimators (Newey (1985); Lee

(1996); Vytlacil (2002); Vytlacil & Yildiz (2007)). However, these estimators are quite difficult to implement in the case where both the corner solution and binary endogeneity occur. Moreover, in a panel data framework, the semiparametric fixed-effect approach cannot identify average partial effects.

Angrist (2001) discussed other alternative methods for estimating dummy endogenous variables (including 2SLS, IV for an exponential conditional mean, minimum mean squared error approximation or quantile treatment effects approach). He prefers IV an estimation strategy (similar to Mullahy (1997) and Abadie (2000)) for nonlinear models with covariates and a nonstructural approach since it gives similar average treatment effects. However, he did not give any evidence for not using the structural approach.

In this chapter, I focus on the simple two-step estimation method (Terza (1998); Kim (2006)) since our model has both corner solution and binary endogeneity problems. It would be attractive to use this method that incorporates a method similar to Heckman (1979) to correct for sample selection. In addition, in our panel data framework, we would like to use correlated random effects to handle heterogeneity in the presence of endogeneity and correlated participation (similar to Semykina & Wooldridge (2010)). Both IV strategies and the bivariate probit method are utilized to handle the binary endogeneity.

The proposed method is very applicable in many economic models since switching endogeneity is of interest to many applied economists and policy makers. One interesting application is estimating the effect of having a newborn on women's labor supply in the presence of a corner solution response and unobserved heterogeneity. Hence, the following part will consist of a literature review on the relationship between fertility and female labor supply.

A remarkable number of studies have examined the effect of fertility on female labor supply and labor force participation. These studies can be divided into four major groups, depending on how they handle the endogeneity problem of the fertility decision. The first group is presented by the studies of Gronau (1973), Heckman (1974), and Heckman & Willis (1977) who assumed exogenous fertility and established a strong negative correlation between female labor supply and fertility. However, as Browning (1992) commented, very few credible inferences can be drawn from them even though we have a number of robust correlations. Their main methodology is to use OLS to estimate the effects of fertility on labor supply.

A second group of studies led by Cain & Dooley (1976), Schultz (1978), and Fleisher & Rhodes (1979) acknowledged endogenous fertility. They handled the endogeneity problem by estimating simultaneous equations models. Smaller estimates on fertility are found when treating it as an endogenous variable than when treating it as an exogenous variable. The problem with this approach is that it is hard to find plausible exclusion restrictions that could identify the underlying structural parameters.

A third group of studies, pioneered by the work of Nakamura & Nakamura (1992), added the lagged dependent variable (i.e. hours of work) to control for unobserved heterogeneity across women. This approach has been used subsequently by a number of authors (Even (1987); Lehrer (1992)). Although adding the lagged dependent variable can help control for unobserved heterogeneity, it still does not address the problem of the endogeneity of the fertility decision.

Last but not least, a fourth group of studies solved the endogeneity problem of fertility by exploiting exogenous sources of variation in family size. Rosenzweig & Wolpin (1980) first used this strategy by comparing the labor supply of women who had twins at their first birth with that of women who had a single child. Then Bronars & Grogger (2001); Jacobsen et al. (1999) used the same strategy but managed to obtain more precise estimates. Other studies (Bloom et al. (2009); Kim & Aassve (2006)) exploit abortion legislation or the contraceptive choice of couples as an IV for fertility. In the same spirit as the twins studies mentioned above, Angrist & Evans (1998) estimated the effect of a third or higher order child on female labor supply by exploiting the fact that parents typically prefer mixed-sex siblings. For a sample of couples with at least two children, they instrumented further childbearing (i.e. having more than two children) with a dummy variable for whether the sex of the second child matched the sex of the first. Because sex mix is virtually random, this strategy allows for identification of the effect of a third or higher order child.

Nguyen (2010) emphasized the negative significant impact of the number of children on female

labor supply. The paper found diminishing impacts of having children on female labor supply and the first child always has the largest adverse effect on a mother's labor supply. This implies that children do not have equal impacts on a mother's labor supply. This finding is similar to the idea from Browning (1992) that having a newborn has more significant impact on a mother's labor supply than having a general number of kids. However, the paper does not view the problem in terms of a two-part model acknowledging the fact that people who decide to work will have positive working hours. This chapter will consider the issue of female labor force participation in a relation with female labor supply and the impact of having a newborn on both a mother's amount and participation decision, which calls again for a discrete endogeneity of having children.

3.3 Model and Estimation

I consider a panel data model with a corner solution response and a binary endogenous variable in the presence of correlated participation decision and heterogeneity as follows:

$$y_{1it} = y_{2it} \exp(X_{1it}\beta_1 + y_{3it}\alpha_1 + c_{1i} + u_{1it}), \qquad (3.1)$$

or

$$\log(y_{1it}) = X_{1it}\beta_1 + y_{3it}\alpha_1 + c_{1i} + u_{1it} \text{ if } y_{1it} > 0 \text{ or } y_{2it} = 1(\text{iff } y_{2it}^* > 0),$$

$$y_{2it}^{r} = X_{2it}\beta_2 + y_{3it}\alpha_2 + c_{2i} + u_{2it}, \qquad (3.2)$$

$$y_{3it}^* = X_{3it}\beta_3 + c_{3i} + u_{3it}, \tag{3.3}$$

$$y_{2it} = 1[y_{2it}^* > 0], (3.4)$$

$$y_{3it} = 1[y_{3it}^* > 0], \tag{3.5}$$

where y_{1it} is continuous with strictly positive values when $y_{2it}^* > 0$ and equal to zero when $y_{2it}^* < 0$ with positive probability, hereafter i = 1, 2, ..., N and t = 1, 2, ..., T. We assume that we observe y_{1it} only when $y_{2it}^* > 0$ or $y_{2it} = 1$. X_{mit} are $1 \times K_m$ vectors of exogenous explanatory variables (for m = 1, 2, 3) which can contain a constant term. β_m are $K_m \times 1$ vectors of parameters. c_{mi} are time-constant unobserved heterogeneity and u_{mit} are idiosyncratic errors. α_1 and α_2 are scalar parameters. 1[·] is an indicator function which has a value of one when the expression inside the bracket is true, otherwise has a value of zero. Both y_{2it} and y_{3it} are dummy variables. We assume a balanced panel for simplicity so i = 1, 2, ..., N and t = 1, 2, ..., T is assumed throughout this chapter. A novel feature of this panel data model is that the common endogenous variable y_{3it} appears in both the amount and participation equations: (3.1) and (3.2). We therefore need to handle both endogeneity and the corner solution problem in equation (3.1).

Following the work of Heckman et al. (1999) and Heckman (1979), α_1 and α_2 are identified if X_{3i} includes at least one variable which is excluded from X_{2i} or X_{1i} under the correct assumption of joint distribution of the error terms. That variable is usually referred to as (an) instrumental variable(s). X_{2i} should include at least one variable which is not in X_{1i} . Those instrumental variables are assumed strictly exogenous conditional on unobserved heterogeneity. With that in mind, and using the modeling device in Mundlak (1978), we can model the relationship between unobserved effects c_{mi} and X_{mit} for each *m*. Let us rewrite equations (3.1)-(3.3) as follows:

$$y_{1it} = y_{2it} \exp(X_{1it}\beta_1 + y_{3it}\alpha_1 + c_{1i} + u_{1it}),$$
(3.6)

$$y_{2it}^* = X_{1it}\beta_{21} + X_{22it}\beta_{22} + y_{3it}\alpha_2 + c_{2i} + u_{2it},$$
(3.7)

$$y_{3it}^* = X_{1it}\beta_{31} + X_{32it}\beta_{32} + c_{3i} + u_{3it},$$
(3.8)

where X_{32it} and X_{22it} are instrumental variables.

Now we assume that:

$$c_{mi} = \overline{Z_i} \delta_m + a_{mi}, \quad m = 1, 2, 3, \tag{3.9}$$

where $a_{mi}|X_{mi} \sim Normal(0, \sigma_{am}^2)$; $\overline{Z}_i = T^{-1} \sum_{t=1}^T Z_{it}$; Z_{it} contains both explanatory variables X_{1it} , X_{32it} , and X_{22it} . \overline{Z}_i is a $1 \times L$ vector where $L = K_2 + K_3 - K_1$.

And now we can rewrite equations (3.1)-(3.3) as:

$$\log(y_{1it}) = W_{1it}\gamma_1 + y_{3it}\alpha_1 + v_{1it} \text{ if } y_{1it} > 0 \text{ or } y_{2it} = 1; W_{1it} \equiv (X_{1it}, \overline{Z_i}),$$
(3.10)

$$y_{2it}^* = W_{2it}\gamma_2 + y_{3it}\alpha_2 + v_{2it}; W_{2it} = (X_{2it}, \overline{Z_i}),$$
(3.11)

$$y_{3it}^* = W_{3it}\gamma_3 + v_{3it}; W_{3it} = (X_{3it}, \overline{Z_i}),$$
(3.12)

where $v_{mit} = a_{mi} + u_{mit}$; m = 1, 2, 3.

As discussed in (Wooldridge, 2010, section 17.6.3), we model the corner solution using the ET2T model. First, we can ensure that the predicted value of the response variable is positive. Second, it is noticeable that we can allow a correlation between unobserved factors that affect the amount equation and unobserved factors that affect the participation equation, that is, v_{1it} and v_{2it} are correlated. This assumption is exploited to relax the assumption in the usual lognormal hurdle model. Moreover, it is a reasonable assumption in empirical study. For example, in the model for married women's labor supply, unobserved factors can influence both women's LFP and labor supply or the unobserved effects determining both decisions are related. Therefore, we can assume that:

$$E(v_{1it}|W_{1i}, v_{2it}) = \eta v_{2it}, \qquad (3.13)$$

in addition to $Var(v_{2it}) = 1$; $Var(v_{1it}) = \sigma^2$; $Cov(v_{1it}, v_{2it}) = \psi \sigma = \eta$ where ψ is the correlation between v_{1it} and v_{2it} .

We are interested in deriving $E(\log(y_{1it})|W_i, y_{3it}, y_{2it}^* > 0)$.

$$E(\log(y_{1it})|W_i, y_{3it}, y_{2it}^* > 0) = W_{1it}\gamma_1 + y_{3it}\alpha_1 + E(v_{1it}|W_i, y_{3it}, y_{2it}^* > 0).$$
(3.14)

Now,

$$E(v_{1it}|W_i, y_{3it}, y_{2it}^* > 0) = y_{3it}E(v_{1it}|W_i, y_{3it}^* > 0, y_{2it}^* > 0) + (1 - y_{3it})E(v_{1it}|W_i, y_{3it}^* < 0, y_{2it}^* > 0),$$

or

$$E(v_{1it}|W_i, y_{3it}, y_{2it}^* > 0) = y_{3it}E_1 + (1 - y_{3it})E_0$$

We will derive E_1 first and apply the similar strategy for E_0 .

$$E_1 = E(v_{1it}|W_i, y_{3it}^* > 0, y_{2it}^* > 0) = \eta_1 E(v_{2it}|W_i, y_{3it} = 1, y_{2it} = 1),$$
(3.15)

$$E_1 = \eta_{12} E_{12} + \eta_{13} E_{13}, \tag{3.16}$$

where

$$E_{12} = \phi(W_{3it}\gamma_3)\Phi\left[(W_{2it}\gamma_2 + \alpha_2 - \rho W_{3it}\gamma_3)(1 - \rho^2)^{-1/2}\right]\Phi_2^{-1}(W_{3it}\gamma_3, W_{2it}\gamma_2 + \alpha_2; \rho), \quad (3.17)$$

and

$$E_{13} = \phi(W_{2it}\gamma_2 + \alpha_2)\Phi\left[(W_{3it}\gamma_3 - \rho(W_{2it}\gamma_2 + \alpha_2))(1 - \rho^2)^{-1/2}\right]\Phi_2^{-1}(W_{3it}\gamma_3, W_{2it}\gamma_2 + \alpha_2; \rho),$$
(3.18)

under the assumption that $Cov(v_{2it}, v_{3it}) = \rho$ and $Var(v_{3it}) = 1$; we can also write: $v_{2it} = \rho v_{3it} + e_{it}$ where $e_{it}|Z_i, v_{3it} \sim Normal(0, \sigma_e^2)$.

$$E_0 = \eta_{02} E_{02} + \eta_{03} E_{03}, \tag{3.19}$$

where

$$E_{02} = \phi(-W_{3it}\gamma_3)\Phi\left[(W_{2it}\gamma_2 - \rho W_{3it}\gamma_3)(1-\rho^2)^{-1/2}\right]\Phi^{-1}(-W_{3it}\gamma_3, W_{2it}\gamma_2, -\rho), \quad (3.20)$$

and

$$E_{03} = \phi(W_{2it}\gamma_2)\Phi\left[(-W_{3it}\gamma_3 + \rho W_{2it}\gamma_2)(1-\rho^2)^{-1/2}\right]\Phi^{-1}(-W_{3it}\gamma_3, W_{2it}\gamma_2, -\rho).$$
(3.21)

Using these two regimes of y_{3it} and correlated participation, we can handle both endogenous switching and corner solution problems. Instead of proceeding with the full information maximum likelihood, we estimate parameters of interest using a two-step estimation procedure based on Heckman's idea of correcting a selection problem using the correction functions (which is the inverse Mill's ratio in Heckman's model). For each regime corresponding to either $y_{3it} = 0$ or $y_{3it} = 1$, we add two correction terms which comprise one part for fixing an endogeneity problem and the other part for correcting correlated unobserved effects bias from the participation equation. The conditional mean of interest with positive outcome, $y_{2it} = 1$, is:

$$E(\log(y_{1it})|W_i, y_{3it}, y_{2it}^* > 0) = \begin{bmatrix} W_{1it}\gamma_1 + y_{3it}\alpha_1 + \eta_{12}y_{3it}E_{12}(\theta_1) + \eta_{13}y_{3it}E_{13}(\theta_1) \\ + (1 - y_{3it})\eta_{02}E_{02}(\theta_1) + \eta_{03}(1 - y_{3it})E_{03}(\theta_1) \end{bmatrix}, \quad (3.22)$$

where $W_i = W_{1i} \cup W_{2i} \cup W_{3i}$ and 4 correction terms E_{12}, E_{13}, E_{02} and E_{03} are stated as above.

We can identify $\theta_1 = (\alpha_2, \gamma'_3, \gamma'_2, \rho)'$, a $Q \times 1$ vector, $(Q = 1 + K_2 + K_3 + 2L + T)$ using maximum likelihood estimation for pooled bivariate probit model in the first stage. Similar to Heckman
(1978a), Greene (1997), and Carrasco (2001), the log likelihood function in the first stage that solves for estimates of θ_1 is:

$$\ln L_{it}(\theta_1) = \begin{bmatrix} (1 - y_{3it})(1 - y_{2it})\ln P_{00} + y_{3it}(1 - y_{2it})\ln P_{10} \\ + y_{2it}(1 - y_{3it})\ln P_{01} + y_{3it}y_{2it}\ln P_{11} \end{bmatrix},$$
(3.23)

where

$$P_{11} = \Pr(y_{3it} = 1 \text{ and } y_{2it} = 1) = \Phi_2(W_{3it}\gamma_3, W_{2it}\gamma_2 + \alpha_2; \rho), \quad (3.24)$$

$$P_{00} = \Pr(y_{3it} = 0 \text{ and } y_{2it} = 0) = \Phi_2(-W_{3it}\gamma_3, -W_{2it}\gamma_2, \rho), \qquad (3.25)$$

$$P_{10} = \Pr(y_{3it} = 1 \text{ and } y_{2it} = 0) = \Phi_2(W_{3it}\gamma_3, -W_{2it}\gamma_2 - \alpha_2; -\rho), \quad (3.26)$$

$$P_{01} = \Pr(y_{3it} = 0 \text{ and } y_{2it} = 1) = \Phi_2(-W_{3it}\gamma_3, W_{2it}\gamma_2; -\rho).$$
(3.27)

We can estimate parameters in the first stage to obtain $\hat{\theta}_1$ (and its standard errors as shown in the first-stage technicalities of Appendix F) and get 4 correction terms in equation (3.22) to plug in the second stage. In the second stage, we estimate the following equation on the selected sample with $y_{2it} = 1$ or a positive dependent variable using POLS:

$$\log(y_{1it}) = \begin{bmatrix} W_{1it}\gamma_1 + y_{3it}\alpha_1 + \eta_{12}y_{3it}E_{12}(\hat{\theta}_1) + \eta_{13}y_{3it}E_{13}(\hat{\theta}_1) \\ + (1 - y_{3it})\eta_{02}E_{02}(\hat{\theta}_1) + \eta_{03}(1 - y_{3it})E_{03}(\hat{\theta}_1) + \varepsilon_{it} \end{bmatrix}.$$
 (3.28)

Hence, with a similar idea of adding the inverse Mills ratio to correct for the sample selection bias, we can add 4 correction terms to control for a corner solution problem with a correlated participation decision and binary endogeneity. We can rewrite the estimating equation above as:

$$\log(y_{1it}) = W_{1it}\gamma_1 + y_{3it}\alpha_1 + \sum_{j=1}^4 \eta_j \hat{\lambda}_{it\,j} + \varepsilon_{it}, \qquad (3.29)$$

where $\hat{\lambda}_{it1} = y_{3it}E_{12}(\hat{\theta}_1)$; $\hat{\lambda}_{it2} = y_{3it}E_{13}(\hat{\theta}_1)$; $\hat{\lambda}_{it3} = (1 - y_{3it})E_{02}(\hat{\theta}_1)$; and $\hat{\lambda}_{it4} = (1 - y_{3it})E_{03}(\hat{\theta}_1)$.

Even though the two-step estimator is easy to implement and numerically robust, we need to adjust the second-stage standard errors, taking into account the first-stage estimation. I show how to obtain $\hat{\theta}_1$ and derive the asymptotic variance of this two-step estimator ($\hat{\theta}_2$) in the technical section of Appendix F.

3.4 Average Partial Effect

The quantity of interest in this study is average treatment effect (ATE) of the binary endogenous variable. We can also obtain average partial effects (APEs) for exogenous explanatory variables.

First we rewrite model (3.1) to (3.3) in the conditional mean forms as follows:

$$E(\log(y_{1it}) = X_{1it}\beta_1 + y_{3it}\alpha_1 + c_{1i} + u_{1it} \text{ if } y_{1it} > 0, \qquad (3.30)$$

$$E(y_{2it}|X_{2it}, y_{3it}, c_{2i}, u_{2it}) = \Phi(X_{2it}\beta_2 + y_{3it}\alpha_2 + c_{2i} + u_{2it}),$$
(3.31)

$$E(y_{3it}|X_{3it},c_{3i},u_{3it}) = \Phi(X_{3it}\beta_3 + c_{3i} + u_{3it}).$$
(3.32)

Our main interest lies in the treatment effect of a binary endogenous variable in both equations (3.30) and (3.31). We can evaluate the effect at values of exogenous explanatory variables of interest. But first, we need to handle the correlated unobserved effects using Mundlak's device as shown in equation (3.9) and follow the estimation procedure that is clarified in the previous section. Now (3.30) and (3.31) have been previously derived as:

$$E(\log(y_{1it})|Z_i, y_{3it}, y_{1it} > 0) = X_{1it}\beta_1 + y_{3it}\alpha_1 + \overline{Z_i}\delta_1 + \sum_{j=1}^4 \eta_j \lambda_{itj},$$
(3.33)

$$E(y_{2it}|Z_i, y_{3it}, a_{2i}, u_{2it}) = \Phi(X_{2it}\beta_2 + y_{3it}\alpha_2 + \overline{Z_i}\delta_2 + a_{2i} + u_{2it}).$$
(3.34)

ATE for the amount equation:

For y_{3t} as a binary variable, the ATE at time *t* can be obtained by averaging equation (3.30) over the distribution of c_{1i} and u_{1it} or take a difference in:

$$E_{(\overline{Z_i},\lambda_{it})}[X_{1t}\beta_1 + y_{3t}\alpha_1 + \overline{Z_i}\delta_1 + \sum_{j=1}^4 \eta_j \lambda_{itj}], \qquad (3.35)$$

where in the argument of the expectation operator, variables with a subscript *i* are random and all others are fixed.

With the definition from equation (3.17) - equation (3.21), plus equation (3.29), (3.35) is rewritten as:

$$E_{E_{it}}[\alpha_1 + \eta_{12}E_{12}(\theta_1) + \eta_{13}E_{13}(\theta_1) - \eta_{02}E_{02}(\theta_1) - \eta_{03}E_{03}(\theta_1)].$$
(3.36)

Given consistent estimator of $\hat{\theta}_1$ and $\hat{\theta}_2$, the ATE of the binary variable y_{3t} in equation (3.33) can be estimated as:

$$\widehat{ATE} = N^{-1} \sum_{i=1}^{N} \left[\widehat{\alpha}_{1} + \widehat{\eta}_{12} E_{12}(\widehat{\theta}_{1}) + \widehat{\eta}_{13} E_{13}(\widehat{\theta}_{1}) - \widehat{\eta}_{02} E_{02}(\widehat{\theta}_{1}) - \widehat{\eta}_{03} E_{03}(\widehat{\theta}_{1}) \right], \quad (3.37)$$

where for each unit we predict the difference in mean responses with and without "treatment" (for $y_{3t} = 1$ and $y_{3t} = 0$), and then average the difference in these estimated mean responses across all units.

ATE for the participation equation:

We rewrite model (3.34) with the scaled coefficients using a standard mixing property of the normal distribution of e_{it} :

$$E(y_{2it}|Z_i, y_{3it}, v_{3it}, e_{it}) = \Phi(X_{2it}\beta_2 + y_{3it}\alpha_2 + \overline{Z_i}\delta_2 + \rho v_{3it} + e_{it}),$$
(3.38)

or

$$E(y_{2it}|Z_i, y_{3it}, v_{3it}, e_{it}) = \Phi(X_{2it}\beta_{2e} + y_{3it}\alpha_{2e} + \overline{Z_i}\delta_{2e} + \rho_e v_{3it}),$$
(3.39)

where the subscript *e* denotes division by $\sqrt{1 + \sigma_e^2}$.

Note that we can write (3.38) - equation (3.39) in terms of bivariate probit model as in the technical section and the procedure to obtain APE or ATE is the same as described below. That means we average out $\overline{Z_i}$ and then take derivatives or changes with respect to the elements of (X_{2t}, y_{3t}) .

The APEs are obtained by computing derivatives, or obtaining differences, in:

$$E_{(\overline{Z_i}, v_{3it})}[\Phi(X_{2t}\beta_{2e} + y_{3t}\alpha_{2e} + \overline{Z_i}\delta_{2e} + \rho_e v_{3it})], \qquad (3.40)$$

or

$$E_{(\overline{Z_i})}[\Phi(X_{2t}\beta_{2\nu} + y_{3t}\alpha_{2\nu} + \overline{Z_i}\delta_{2\nu})], \qquad (3.41)$$

where the subscript *v* denotes division by $\rho_e \sqrt{1 + \sigma_{v_3}^2}$ and $Var(v_{3it}) = \sigma_{v_3}^2 = 1$. In order to obtain partial effects, we average out $\overline{Z_i}$ and then take derivatives or changes with respect to the elements

of (X_{2t}, y_{3t}) . Across the sample for a chosen *t*, we can obtain the estimators for APE with respect to one element X_{2t1} of X_{2t} as:

$$\widehat{APE} = \widehat{\beta}_{2\nu 1} \left[N^{-1} \sum_{i=1}^{N} \phi(X_{2t} \widehat{\beta}_{2\nu} + y_{3t} \widehat{\alpha}_{2\nu} + \overline{Z_i} \widehat{\delta}_{2\nu}) \right].$$
(3.42)

The estimator for ATE with respect to y_{3t} is:

$$\widehat{ATE} = N^{-1} \sum_{i=1}^{N} \left[\Phi(X_{2t} \widehat{\beta}_{2\nu} + \widehat{\alpha}_{2\nu} + \overline{Z_i} \widehat{\delta}_{2\nu}) - \Phi(X_{2t} \widehat{\beta}_{2\nu} + \overline{Z_i} \widehat{\delta}_{2\nu}) \right],$$
(3.43)

which we are interested in.

3.5 Empirical Example

3.5.1 Overview of Data

Over the past two decades, fertility has decreased as the labor force participation rates of women in most developing and advanced countries have increased (Kim & Aassve (2006)). This change implies the changing roles of women and changes in the time allocation among household members in both work activities and fertility behavior. We also observed this pattern in Vietnam.

For the last two decades, the fertility rates of Vietnamese women fell while the labor force participation rates for the whole population did not change very much. A decline in fertility also accompanied an increase in income. During the period from 1986 to 2006, while fertility dramatically decreased, GDP per capita increased 2.9 times to 587.4 USD per capita. This pattern is consistent with microeconomic predictions: higher income leads to a reduction in fertility and the inverse relationship of fertility and labor force participation (Becker & Lewis (1973); Willis (1973)). Thus, it is important to analyze the data on fertility and labor market behavior of working women.

The data used in this paper came from the Vietnamese Household Living Standard Surveys (VHLSS) 2004, 2006, and 2008, which were conducted by the Vietnamese General Statistical Office (GSO) with technical support from The World Bank. The survey sample was randomly

selected to represent the whole country, taking into account urban and rural structures, geographical conditions, regional issues, ethnic differences, and provincial representation. The sample used in this chapter has 665 women. The survey collected information about the following: household information, education, health, employment, migration, housing, fertility and family planning, incomes, expenditures, borrowing, lending, and savings.

Only households with children under 18 years old and households with a mother and father younger than 60 and 65 years of age, respectively, at the time of the interview are included in this research. There are 1,995 households in the sample used for this research. Table C.1 provides a summary of the descriptive statistics for the whole sample. The dependent variables are working status and hours worked per day for a woman (being either head of household or spouse). According to Table C.1, 95% of mothers worked in the interview year, and on average, they worked 7.8 hours per day. The explanatory variables are whether mother has a newborn, mother's education, age, non-labor income, father's education, age; and other household characteristics such as whether they live in an urban area, they work on a farm and their ethnicity. In this sample, each household had an average of 2.5 children; 10% of the sample women had newborns; 55% and 56% women in the sample had a boy first and their first two kids had the same gender, respectively. In general, the husband's education is higher than wife's education. Income from other sources for women in the sample is about 8 million VND per year (approximately 400-450 USD per year). Table C.1 also shows that around 84% of working wives worked on farms and 18% of households were located in urban areas.

Table C.2 shows the summary statistics for each year in the panel data. There is no obvious pattern for women working hours and labor force participation (LFP). However, we can observe that the fertility rate declines over time. The percentage of having a newborn goes down from 16% in 2004 to 9% in 2006 and 6% in 2008. On the contrary, non-wife income increases over time.

3.5.2 Estimation and Result

The main contribution of this chapter and its following application is to allow the correlation between women's decision to participate in the labor market and their amount of working hours; to acknowledge the nature of having a newborn as a dummy variable and to consider the influence of having a newborn on both women's participation and labor supply. As shown in the literature, newborns have negative effect on women's labor force entry. This means that women who are not working are unlikely to take part in the labor market after delivering babies. This raises the question of how newborns affect their mothers' labor supply for those women who are participating and stay in the market.

When a mother has a newborn, she will decide how many working hours she will spend after deliver a baby. If endogeneity of fertility is not accounted for, we will not obtain consistent estimates of labor supply conditional on fertility. In order to draw robust and credible estimates of the effects of newborns on women's labor supply and participation, we need to take into account this endogeneity. Another important point is both amount and participation decisions are jointly determined because preferences for working or work time somehow are positively correlated. In the same way, preferences for having a baby and for working are negatively correlated. Therefore, we should model these decisions with a joint relationship. Using the panel data VHLSSs 2004-2008, we study women's labor supply in a system of equations where fertility decision is an endogenous dummy variable occurred in both labor supply and participation equations, and this system of equations are jointly correlated. We are interested in estimating a panel data model of working hours for a woman i at time t, who takes having a newborn into consideration as an endogenous factor, as follows:

$$Log(Hours_{it}) = \begin{bmatrix} Newborn_{it}\alpha_1 + Medu_{it}\beta_{11} + Mage_{it}\beta_{12} + Magesq_{it}\beta_{13} \\ + NMincome_{it}\beta_{14} + c_{1i} + u_{1it} \text{ if } Hours_{it} > 0 \end{bmatrix}, \quad (3.44)$$

٦

г

$$LFP_{it}^{*} = \begin{bmatrix} Medu_{it}\beta_{21} + Mage_{it}\beta_{22} + Magesq_{it}\beta_{23} + Hedu_{it}\beta_{24} + Hage_{it}\beta_{25} + \\ Hagesq_{it}\beta_{26} + NMincome_{it}\beta_{27} + Newborn_{it}\alpha_{2} + c_{2i} + u_{2it} \end{bmatrix}, \quad (3.45)$$

$$Newborn_{it}^{*} = \begin{bmatrix} Samesex\beta_{32} + Medug_{it}\beta_{33} + Mage_{it}\beta_{34} \\ + Magesq_{it}\beta_{35} + NMincome_{it}\beta_{36} + c_{3i} + u_{3it} \end{bmatrix}.$$
(3.46)

These equations correspond to equation (3.1) to equation (3.3) in the model section. *Hours*_{it} is annual working hours for a woman *i* at time *t*, which is determined by her education, *Medu*, her age, *Mage*, her age square, *Magesq*, her other income not from her wage, *NMincome*, other variables such as whether she lives in an urban area, whether her ethnicity is majority, whether she works on a farm, and whether she has a newborn, *Newborn*, with the age from 0 to 1. A woman's LFP is influenced by her characteristics (the same variables in equation (3.44), her husband's characteristics including education, age, age square, *Hedu*, *Hage*, *Hagesq*, non-mom income as well as whether she has a newborn. The fertility decision equation has right-hand-side variables including an instrumental variable: whether the first two children have the same gender and other exogenous variables including mom's characteristics and non-mom income. We also allow some explanatory variables to be correlated with heterogeneity and take care of this relationship by adding time averages of explanatory variables into each equation.

With the new procedure to control for a corner solution, we can allow unobserved factors that affect both amount and participation equations to be correlated. In addition, to ensure that the predicted value of labor supply is positive, we need to apply Type II Tobit model to log(hours) rather than hours. That is why we use the specification of Exponential Type II Tobit model (ET2T) (see more in (Wooldridge, 2010, Chapter 17)). In addition, the ET2T model is applicable when we have exclusion restrictions. The participation equation contains many more variables which are not in the amount equation so that the parameters in the amount equation will be identified.

The choice of appropriate instrumental variables is important because these can affect the reliability of estimates and inferences. Valid and strong instrumental variables must satisfy two conditions: an instrumental variable should be uncorrelated with the error term and it should be highly correlated with the right-hand-side endogenous regressor(s). In this research, that means that the instrumental variables have no correlation with factors that directly affect parental LFP and labor supply and that the instruments are correlated with fertility. Whether the first two children have the same gender is used to generate exogenous variations in fertility in this research.

Normally, the gender of a child is a random variable, and it is uncorrelated with parental LFP and labor supply. In addition, we found that the boy-to-girl ratio of the first child was 1.05 in our sample, which is close to the natural ratio. Thus, the gender of the first child is a valid instrumental variable. However, this instrumental variable is not significant in the first stage. Angrist & Evans (1998) found that parents prefer a mixed sibling-sex composition, and parents who first had two girls or boys had a higher probability of having additional children. Carrasco (2001) also found same sex instrumental variable is a strong instrument in the US data. In this dataset, among women with more than two children, the likelihood of another birth was 28% if they had a son, 34% if they had two sons and 11% if they had three sons. This evidence implies that siblings with mixed genders are desirable among Vietnamese families. The same gender of the first two children variables meets the two conditions required of a valid and strong instrument, and it can serve as an instrumental variable to generate exogenous variations in fertility. The same gender of the first two children equals 1 if the first two children have the same gender, and 0 otherwise. According to Table C.1, 55% of sampled households had a male first child and 56% of households had the first two children with the same gender. The t-test is implemented to see if the same gender instrument is strong or not. The result is -3.2, implying that this instrument can be used for this study.

Table C.3 shows the estimation result for the bivariate probit model in the first stage. The coefficient on samesex is positive and it is statistically significant implying samesex is a good and significant instrument in our study. The coefficient on a newborn in the LFP equation is also negative and statistically significant. The effect of a newborn reduces the mother's probability of LFP by 13.6%. In terms of the average treatment effect, compared to women without newborn babies, mothers with newborns have lower probability to continue to work by 12.7%. The coefficient on ρ , -0.165, shows us that there is a negative correlation between unobserved effects that affect both fertility and women's LFP. This brings more evidence to empirical studies of developing countries that having an additional child will negatively influence the probability of working women who just delivered a child to come back to work. Table C.4 reports the coefficient estimates from six different estimation methods. Pooled OLS (POLS) assumes that all explanatory variables are uncorrelated with unobserved heterogeneity and are also strictly exogenous. The estimates based on POLS show that having a newborn reduces the mother's working hours by 13.4%. The POLS estimates have the largest bias because they do not take into account endogeneity of fertility, the presence of heterogeneity which might be correlated with explanatory variables, and the correlation between work participation and the amount of work.

Pooled 2SLS takes into account endogeneity of a newborn but does not remove an unobserved effect. Controlling for endogeneity of a newborn reduces the bias by 10%. Now having a newborn will make a mother reduce her working hours by 23%. Fixed effects (FE) allows for correlation between the explanatory variables and unobserved heterogeneity and FE-2SLS further allows a newborn to be correlated with the idiosyncratic errors. Columns (3) and (4) show that mothers' working hours are diminished by 16.4% and 27.7% using the FE and FE-2SLS. However, FE-2SLS ignores the correlation between women's decision to participate and how much to work. In addition, all methods from (1) to (4) do not consider a newborn a dummy variable.

To take into account the correlated participation, we can also use Heckman type IV correction (see Semykina & Wooldridge (2010)) method and hereafter, we call this estimator SW (under column 5 of Table C.4). This estimator allows correlated participation and heterogeneity in the presence of endogeneity. However, this method ignores the binary nature of the endogeneity and assumes a linear reduced form (using pooled 2SLS in the second stage after obtaining the inverse Mills ratio in the first stage). The result shows that mothers' working hours are reduced by 30.8% using the SW, which is more than the reduction in mothers' working hours using the FE and FE-2SLS. It suggests that correlated participation does matter. However, this decrease is still smaller than the reduction in mothers' working hours using the new proposed procedure since we need to take care of the binary endogeneity.

The new proposed procedure corrects for endogeneity of a newborn, plus its dummy nature and its influence on both women's participation and amount of work. It also reduces another source of bias from correlated heterogeneity by adding time averages of explanatory variables into all equations and time dummies. However, the standard errors are larger once these corrections are accounted for. After controlling for all these sources of bias, women who are still working will decrease their working hours by 34.5%.

The result shows that having a new child in Vietnamese households has a negative effect on maternal hours for working women. Women will have to give up their working hours by 34.5% to take care of her newborn or use such forgone time as an input of home production.

3.6 Conclusion

This chapter studies the nonlinear panel data model with an endogenous dummy variable and a corner solution response. The main contribution is to allow a joint distribution of the endogenous dummy regressor and unobserved factors that affect both the amount and participation equations. I propose a two-step estimation method in which the first stage exploits a bivariate probit model for the relationship between the endogenous dummy variable and the participation decision. For the amount equation, by using an ET2T model, we can ensure that the predicted value of log(hours) is positive; and there is a correlation between unobserved effects in both the amount and participation equations. In addition, we need to allow exclusion restrictions in order to identify the parameters in the amount equation. In other words, we allow a set of explanatory variables in the participation equations. I also allow some explanatory variables to be correlated with heterogeneity.

This estimation method is applied to analyze the effect of fertility on women's working hours and labor force participation. The proposed approach gives a statistically significant negative effect of having a newborn on a woman who is working and remain in the labor market. Having a newborn has a significant negative impact on a woman's taking part into the labor force and her working hours. The proposed estimation method corrects remarkably the bias in estimating the effect of a newborn on a mother's working hours compared to other alternative estimation methods. APPENDICES

Appendix A

TABLES FOR CHAPTER 1

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y ₂ is a	ssumed exe	ogenous	y_2 is assumed endogenous						
y_2 continuous	-0.2347	-0.1283	-0.1591	-0.2079	-0.1583	-0.1754	-0.2295	-0.2368	-0.2371		
		(0.0046)	(0.0042)	(0.0051)	(0.0110)	(0.0064)	(0.0077)	(0.0051)	(0.0050)		
		[.0034]	[.0024]	[.0008]	[.0024]	[.0019]	[.0002]	[.00008]	[.00008]		
y_2 discrete 0-1	-0.32		-0.2014	-0.2763		-0.2262	-0.3109	-0.3201	-0.3204		
			(0.0046)	(0.0051)		(0.0082)	(0.0099)	(0.0041)	(0.0030)		
			[.0038]	[.0014]		[.0030]	[.0003]	[.00005]	[.00001]		
y_2 discrete 1-2	-0.1273		-0.161	-0.1193		-0.1716	-0.1258	-0.128	-0.1278		
			(0.0027)	(0.0017)		(0.0031)	(0.0023)	(0.0020)	(0.0016)		
			[.0011]	[.0002]		[.0014]	[.00005]	[.00004]	[.00001]		
y_2 discrete 2-3	-0.0212		-0.0388	-0.0259		-0.0317	-0.0224	-0.0214	-0.0212		
			(0.0030)	(0.0012)		(0.0031)	(0.0014)	(0.0014)	(0.0010)		
			[.0006]	[.0001]		[.0003]	[.00004]	[.00001]	[.000001]		
<i>x</i> ₁	0.0235	0.0224	0.021	0.0223	0.0237	0.0212	0.0231	0.024	0.0238		
		(0.0181)	(0.0125)	(0.0130)	(0.0189)	(0.0131)	(0.0142)	(0.0159)	(0.0140)		
<i>x</i> ₂	0.0235	0.0218	0.0214	0.0195	0.023	0.0218	0.0241	0.0243	0.0244		
		(0.0181)	(0.0128)	(0.0129)	(0.0192)	(0.0131)	(0.0134)	(0.0153)	(0.0136)		

Table A.1: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional
	value			Probit		BS	Probit	Probit	Probit
Estimation Method	Coef.	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE
		y_2 is a	ssumed exe	ogenous		y ₂ is	assumed end	ogenous	
У2	-1	-0.1283	-0.2024	-0.8543	-0.1583	-0.2275	-0.9387	-1.045	-1.044
		(0.0044)	(0.0046)	(0.0146)	(0.0089)	(0.0084)	(0.0255)	(0.0483)	(0.0424)
<i>x</i> ₁	0.1	0.0224	0.0267	0.0917	0.0237	0.0275	0.0945	0.1061	0.1052
		(0.0181)	(0.0160)	(0.0534)	(0.0190)	(0.0171)	(0.0578)	(0.0702)	(0.0619)
<i>x</i> ₂	0.1	0.0218	0.0272	0.0956	0.0231	0.0282	0.0987	0.1071	0.1073
		(0.0181)	(0.0163)	(0.0534)	(0.0192)	(0.0170)	(0.0548)	(0.0681)	(0.0600)

Table A.2: Simulation Result of the Coefficient Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)

Note: Figures in parenthesis () are standard deviations.

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y ₂ is a	ssumed ex	ogenous	y_2 is assumed endogenous						
y_2 continuous	-0.2461	-0.1507	-0.1854	-0.2402	-0.16	-0.1887	-0.2442	-0.249	-0.2491		
		(0.0042)	(0.0037)	(0.0046)	(0.0102)	(0.0053)	(0.0056)	(0.0044)	(0.0043)		
		[.0031]	[.0019]	[.0002]	[.0027]	[.0018]	[.0001]	[.0001]	[.0001]		
y_2 discrete 0-1	-0.3383		-0.239	-0.3289		-0.2445	-0.3355	-0.3384	-0.3385		
			(0.0042)	(0.0031)		(0.0066)	(0.0051)	(0.0019)	(0.0020)		
			[.0032]	[.0003]		[.0030]	[.00009]	[.000005]	[.000007]		
y_2 discrete 1-2	-0.1332		-0.2001	-0.1319		-0.2022	-0.1331	-0.1332	-0.1332		
			(0.0018)	(0.0011)		(0.0025)	(0.0013)	(0.0008)	(0.0010)		
			[.0021]	[.00004]		[.0022]	[.000004]	[.000001]	[.000001]		
y_2 discrete 2-3	-0.0208		-0.0193	-0.0219		-0.0177	-0.0212	-0.0208	-0.0208		
			(0.0029)	(0.0007)		(0.0032)	(0.0008)	(0.0007)	(0.0007)		
			[.00005]	[.00003]		[.0001]	[.00001]	[.000001]	[.000001]		
<i>x</i> ₂	0.0246	0.0267	0.021	0.025	0.0265	0.0234	0.025	0.0255	0.0253		
		(0.0168)	(0.0089)	(0.0063)	(0.0170)	(0.0090)	(0.0065)	(0.0072)	(0.0066)		
<i>x</i> ₂	0.0246	0.0241	0.0214	0.0246	0.0242	0.0222	0.0246	0.0252	0.0249		
		(0.0178)	(0.0100)	(0.0070)	(0.0182)	(0.0100)	(0.0070)	(0.0077)	(0.0072)		

Table A.3: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.1$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y_2 is a	ssumed ex	ogenous	y_2 is assumed endogenous						
y_2 continuous	-0.2178	-0.1104	-0.1368	-0.1777	-0.1548	-0.1637	-0.2144	-0.2208	-0.2205		
		(0.0042)	(0.0039)	(0.0054)	(0.0148)	(0.0096)	(0.0124)	(0.0052)	(0.0045)		
		[.0034]	[.0026]	[.0013]	[.0020]	[.0017]	[.0001]	[.0001]	[.0001]		
y_2 discrete 0-1	-0.2973		-0.1706	-0.2307		-0.21	-0.2871	-0.3	-0.2994		
			(0.0049)	(0.0069)		(0.0136)	(0.0196)	(0.0054)	(0.0040)		
			[.0040]	[.0021]		[.0028]	[.0003]	[.00008]	[.00006]		
y_2 discrete 1-2	-0.1281		-0.1303	-0.111		-0.1491	-0.1232	-0.1288	-0.1291		
			(0.0031)	(0.0024)		(0.0060)	(0.0047)	(0.0025)	(0.0020)		
			[.00007]	[.0005]		[.0007]	[.0002]	[.00002]	[.00003]		
y_2 discrete 2-3	-0.0253		-0.0532	-0.0319		-0.0452	-0.0258	-0.0247	-0.0249		
			(0.0022)	(0.0019)		(0.0030)	(0.0023)	(0.0021)	(0.0015)		
			[.0009]	[.0002]		[.0006]	[.00002]	[.00002]	[.00001]		
<i>x</i> ₁	0.0218	0.0327	0.0276	0.0291	0.0263	0.0273	0.0305	0.0318	0.0313		
		(0.0222)	(0.0176)	(0.0182)	(0.0169)	(0.0184)	(0.0201)	(0.0237)	(0.0208)		
<i>x</i> ₂	0.0218	0.0215	0.0212	0.0236	0.0244	0.0199	0.0233	0.02	0.0203		
		(0.0201)	(0.0170)	(0.0179)	(0.0184)	(0.0187)	(0.0206)	(0.0216)	(0.0187)		

Table A.4: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.9$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y ₂ is a	ssumed exe	ogenous	y_2 is assumed endogenous						
y_2 continuous	-0.2402	-0.1352	-0.1618	-0.2094	-0.1661	-0.1823	-0.2327	-0.2405	-0.2407		
		(0.0047)	(0.0037)	(0.0050)	(0.0621)	(0.0363)	(0.0366)	(0.0046)	(0.0043)		
		[.0033]	[.0025]	[.0010]	[.0023]	[.0023]	[.0002]	[.00001]	[.00001]		
y_2 discrete 0-1	-0.3202		-0.1992	-0.2724		-0.2301	-0.3157	-0.3199	-0.3202		
			(0.0044)	(0.0055)		(0.0522)	(0.0799)	(0.0037)	(0.0029)		
			[.0038]	[.0015]		[.0018]	[.0001]	[.00001]	[.000001]		
y_2 discrete 1-2	-0.1275		-0.1605	-0.1195		-0.1676	-0.1248	-0.128	-0.1279		
			(0.0026)	(0.0016)		(0.0189)	(0.0087)	(0.0016)	(0.0015)		
			[.0010]	[.0003]		[.0029]	[.00009]	[.00002]	[.00001]		
y_2 discrete 2-3	-0.0213		-0.0386	-0.0268		-0.0332	-0.0227	-0.0215	-0.0214		
			(0.0027)	(0.0011)		(0.0108)	(0.0066)	(0.0012)	(0.0010)		
			[.0005]	[.0002]		[.0013]	[.00005]	[.000001]	[.000003]		
<i>x</i> ₁	0.024	0.0224	0.0114	0.0117	0.0237	0.0235	0.0253	0.0261	0.0252		
		(0.0181)	(0.0118)	(0.0130)	(0.0189)	(0.0225)	(0.0142)	(0.0146)	(0.0127)		
<i>x</i> ₂	0.024	0.0218	0.0104	0.0109	0.023	0.0227	0.0227	0.0244	0.025		
		(0.0181)	(0.0123)	(0.0129)	(0.0192)	(0.0244)	(0.0134)	(0.0135)	(0.0119)		

Table A.5: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications, $\delta_{23} = 0.3$)

This table presents the case of weak IV ($\delta_{23} = 0.3$).

Model	True	Linear	Tobit	Fractional	Tobit	Fractional	Fractional	Fractional
	value			Probit	BS	Probit	Probit	Probit
Estimation Method	APE	OLS	MLE	QMLE	MLE	QMLE-PW	NLS	QMLE
		y_2 is a	ssumed exe	ogenous		y_2 is assume	d endogenou	S
y_2 continuous	-0.2441	-0.1382	-0.1652	-0.2117	-0.1827	-0.2625	-0.2436	-0.2441
		(0.0045)	(0.0034)	(0.0049)	(0.0424)	(0.0427)	(0.0045)	(0.0044)
		[.0033]	[.0024]	[.0010]	[.0020]	[.0002]	[.00001]	[.00001]
y_2 discrete 0-1	-0.3194		-0.2019	-0.2708	-0.2284	-0.3513	-0.3186	-0.3194
			(0.0039)	(0.0053)	(0.0562)	(0.0931)	(0.0036)	(0.0030)
			[.0038]	[.0014]	[.0018]	[.0001]	[.00001]	[.000001]
y_2 discrete 1-2	-0.1269		-0.1605	-0.1189	-0.1697	-0.1294	-0.1274	-0.127
			(0.0025)	(0.0021)	(0.0211)	(0.0101)	(0.0016)	(0.0015)
			[.0010]	[.0002]	[.0015]	[.00007]	[.00001]	[.00001]
y_2 discrete 2-3	-0.0211		-0.036	-0.0269	-0.0267	-0.0175	-0.0213	-0.021
			(0.0026)	(0.0020)	(0.0129)	(0.0074)	(0.0011)	(0.0010)
			[.0004]	[.0002]	[.0010]	[.00004]	[.000001]	[.000002]

Table A.6: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications, $\delta_{23} = 0$)

This table presents the case of no instrument ($\delta_{23} = 0$).

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y ₂ is a	y_2 is assumed exogenous			y_2 is assumed endogenous					
y_2 continuous	-0.235	-0.1419	-0.1695	-0.218	-0.1688	-0.1837	-0.234	-0.2371	-0.2366		
		(0.0221)	(0.0193)	(0.0216)	(0.0667)	(0.0356)	(0.0253)	(0.0166)	(0.0162)		
		[.094]	[.0066]	[.0017]	[.0066]	[.0051]	[.0001]	[.0002]	[.00017]		
y_2 discrete 0-1	-0.3281		-0.2173	-0.3		-0.2386	-0.3277	-0.3288	-0.3281		
			(0.0253)	(0.0339)		(0.0492)	(0.0457)	(0.0137)	(0.0122)		
			[.0111]	[.0028]		[.0090]	[.00004]	[.00005]	[.00004]		
y_2 discrete 1-2	-0.1308		-0.1767	-0.1252		-0.184	-0.1305	-0.13	-0.1306		
			(0.0222)	(0.0096)		(0.0257)	(0.0139)	(0.0073)	(0.0063)		
			[.0046]	[.0006]		[.0053]	[.00004]	[.00009]	[.00004]		
y_2 discrete 2-3	-0.0214		-0.0316	-0.0243		-0.0286	-0.0217	-0.0211	-0.0213		
			(0.0129)	(0.0034)		(0.0134)	(0.0041)	(0.0036)	(0.0027)		
			[.0010]	[.0003]		[.0007]	[.00003]	[.00003]	[.00001]		

Table A.7: Simulation Result of the Average Partial Effects Estimates (N=100, $\eta_1 = 0.5$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional			
	value			Probit		BS	Probit	Probit	Probit			
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE			
		y_2 is a	y_2 is assumed exogenous			y_2 is assumed endogenous						
y_2 continuous	-0.2358	-0.1415	-0.171	-0.2201	-0.1617	-0.1815	-0.2334	-0.2379	-0.2376			
		(0.0177)	(0.0157)	(0.0163)	(0.0175)	(0.0114)	(0.0119)	(0.0051)	(0.0086)			
		[.0046]	[.0034]	[.0007]	[.0041]	[.0029]	[.0001]	[.0001]	[.0001]			
y_2 discrete 0-1	-0.3285		-0.219	-0.3026		-0.2351	-0.3241	-0.3293	-0.329			
			(0.0219)	(0.0311)		(0.0153)	(0.0192)	(0.0109)	(0.0106)			
			[.0059]	[.0012]		[.0052]	[.0002]	[.00001]	[.00002]			
y_2 discrete 1-2	-0.1309		-0.1782	-0.1259		-0.1847	-0.13	-0.131	-0.1311			
			(0.0205)	(0.0082)		(0.0158)	(0.0058)	(0.0044)	(0.0043)			
			[.0028]	[.0002]		[.0031]	[.00004]	[.00001]	[.000004]			
y_2 discrete 2-3	-0.0214		-0.0309	-0.024		-0.0267	-0.0219	-0.0212	-0.0213			
			(0.0109)	(0.0025)		(0.0084)	(0.0018)	(0.0017)	(0.0014)			
			[.0004]	[.0001]		[.0002]	[.00002]	[.00001]	[.000004]			

Table A.8: Simulation Result of the Average Partial Effects Estimates (N=500, $\eta_1 = 0.5$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional			
	value			Probit		BS	Probit	Probit	Probit			
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE			
		y ₂ is a	y_2 is assumed exogenous			y_2 is assumed endogenous						
y_2 continuous	-0.2347	-0.1283	-0.1591	-0.2079	-0.1583	-0.1754	-0.2295	-0.2368	-0.2371			
		(0.0046)	(0.0042)	(0.0051)	(0.0110)	(0.0064)	(0.0077)	(0.0051)	(0.0050)			
		[.0034]	[.0024]	[.0008]	[.0024]	[.0019]	[.0002]	[.00008]	[.00008]			
y_2 discrete 0-1	-0.32		-0.2014	-0.2763		-0.2262	-0.3109	-0.3201	-0.3204			
			(0.0046)	(0.0051)		(0.0082)	(0.0099)	(0.0041)	(0.0030)			
			[.0038]	[.0014]		[.0030]	[.0003]	[.00016]	[.00001]			
y_2 discrete 1-2	-0.1273		-0.161	-0.1193		-0.1716	-0.1258	-0.128	-0.1278			
			(0.0027)	(0.0017)		(0.0031)	(0.0023)	(0.0020)	(0.0016)			
			[.0011]	[.0002]		[.0014]	[.00005]	[.00004]	[.00001]			
y_2 discrete 2-3	-0.0212		-0.0388	-0.0259		-0.0317	-0.0224	-0.0214	-0.0212			
			(0.0030)	(0.0012)		(0.0031)	(0.0014)	(0.0014)	(0.0010)			
			[.0006]	[.0001]		[.0003]	[.00004]	[.00001]	[.000001]			

Table A.9: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional		
	value			Probit		BS	Probit	Probit	Probit		
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE		
		y ₂ is a	ssumed exe	ogenous	y_2 is assumed endogenous						
y_2 continuous	-0.2347	-0.1286	-0.1591	-0.208	-0.1591	-0.1755	-0.2293	-0.2369	-0.2371		
		(0.0028)	(0.0028)	(0.0031)	(0.0082)	(0.0044)	(0.0050)	(0.0031)	(0.0030)		
		[.0024]	[.0017]	[.0006]	[.0017]	[.0013]	[.0001]	[.00005]	[.00006]		
y_2 discrete 0-1	-0.3201		-0.2014	-0.2766		-0.2263	-0.3106	-0.3201	-0.3204		
			(0.0034)	(0.0036)		(0.0059)	(0.0074)	(0.0029)	(0.0021)		
			[.0027]	[.0010]		[.0021]	[.0002]	[.000001]	[.000007]		
y_2 discrete 1-2	-0.1275		-0.1609	-0.1194		-0.1717	-0.1258	-0.1281	-0.1278		
			(0.0020)	(0.0012)		(0.0024)	(0.0017)	(0.0015)	(0.0011)		
			[.0008]	[.0002]		[.0010]	[.00004]	[.00001]	[.000009]		
y_2 discrete 2-3	-0.0213		-0.039	-0.0259		-0.0317	-0.0224	-0.0214	-0.0212		
			(0.0020)	(0.0008)		(0.0021)	(0.0010)	(0.0010)	(0.0007)		
			[.0004]	[.0001]		[.0002]	[.00003]	[.000002]	[.000001]		

Table A.10: Simulation Result of the Average Partial Effects Estimates (N=2000, $\eta_1 = 0.5$, 500 replications)

Model	True	Linear	Tobit	Fractional	Linear	Tobit	Fractional	Fractional	Fractional			
	value			Probit		BS	Probit	Probit	Probit			
Estimation Method	APE	OLS	MLE	QMLE	2SLS	MLE	QMLE-PW	NLS	QMLE			
		y_2 is a	y_2 is assumed exogenous			y_2 is assumed endogenous						
y_2 continuous	-0.2379	-0.1599	-0.1876	-0.2369	-0.1625	-0.1885	-0.2375	-0.239	-0.239			
		(0.0053)	(0.0041)	(0.0050)	(0.0088)	(0.0051)	(0.0056)	(0.0049)	(0.0048)			
		[.0025]	[.0015]	[.00003]	[.0024]	[.0016]	[.00001]	[.00003]	[.00003]			
y_2 discrete 0-1	-0.3409		-0.2431	-0.3393		-0.2445	-0.3403	-0.3401	-0.3401			
			(0.0040)	(0.0028)		(0.0063)	(0.0050)	(0.0023)	(0.0018)			
			[.0031]	[.00005]		[.0031]	[.00002]	[.00003]	[.00003]			
y_2 discrete 1-2	-0.1361		-0.2032	-0.1358		-0.2037	-0.136	-0.136	-0.136			
			(0.0019)	(0.0010)		(0.0026)	(0.0012)	(0.0011)	(0.0010)			
			[.0021]	[.000007]		[.0021]	[.000002]	[.000002]	[.000002]			
y_2 discrete 2-3	-0.0215		-0.0195	-0.0217		-0.0192	-0.0216	-0.0216	-0.0216			
			(0.0027)	(0.0006)		(0.0032)	(0.0007)	(0.0008)	(0.0006)			
			[.00006]	[.000005]		[.00007]	[.000003]	[.000003]	[.000003]			
<i>x</i> ₁	0.0238	0.0265	0.0223	0.024	0.0237	0.0224	0.024	0.0242	0.024			
		(0.0165)	(0.0084)	(0.0059)	(0.0189)	(0.0084)	(0.0059)	(0.0064)	(0.0061)			
<i>x</i> ₂	0.0238	0.0234	0.0217	0.024	0.023	0.0218	0.024	0.0239	0.0239			
		(0.0179)	(0.0103)	(0.0064)	(0.0192)	(0.0104)	(0.0064)	(0.0064)	(0.0061)			

Table A.11: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, a1 is normally distributed, 500 replications)

Table A.12: Simulation Result of the Average Partial Effects Estimates (N=1000, $\eta_1 = 0.5$, 500 replications)

APE (QMLE)	True APE	Mean	SD	MSE	Rejection rate
y_2 continuous	-0.2347	-0.2371	0.005	0.0051	0.046
y_2 discrete 0-1	-0.32	-0.3204	0.003	0.0029	0.045
y_2 discrete 1-2	-0.1273	-0.1278	0.0016	0.0014	0.046
y_2 discrete 2-3	-0.0212	-0.0212	0.001	0.0009	0.048

Model	Fractional Probit				
Estimation Method	Q	MLE	NLS		
Standard error	analytical	bootstrapping	analytical	bootstrapping	
y_2 continuous	-0.2406	-0.2406	-0.2405	-0.2405	
	(0.0043)	(0.0041)	(0.0046)	(0.0043)	
y_2 discrete 0-1	-0.3203	-0.3203	-0.32	-0.32	
	(0.0030)	(0.0028)	(0.0038)	(0.0034)	
y_2 discrete 1-2	-0.1279	-0.1279	-0.128	-0.128	
	(0.0015)	(0.0013)	(0.0016)	(0.0014)	
y_2 discrete 2-3	-0.0214	-0.0214	-0.0215	-0.0215	
	(0.0010)	(0.0010)	(0.0012)	(0.0012)	

Table A.13: Comparison of analytical and bootstrapping mean of standard errors (N=1000, $\eta_1 = 0.5$, 200 replications)

Note: Figures in parenthesis () are mean of standard errors. Figures not in parenthesis () are APEs' estimates. Bootstrapping standard errors are obtained by bootstrapping method using 100 bootstrap replications.

Number	Frequency	Percent	Cumulative
of kids			relative frequency
0	16,200	50.9	50.9
1	10,000	31.42	82.33
2	3,733	11.73	94.06
3	1,373	4.31	98.37
4	323	1.01	99.39
5	134	0.42	99.81
6	47	0.15	99.96
7	6	0.02	99.97
8	4	0.01	99.99
9	2	0.01	99.99
10	2	0.01	100
Total	31,824	100	

Table A.14: Frequencies of the Number of Children

Table A.15: Descriptive Statistics

Variable	Description	Mean	S.D.	Min	Max
frhour	Women's weekly fractional working hours	0.126	0.116	0	0.589
kidno	Number of kids	0.752	0.977	0	10
age	Mother's age in years	29.742	3.613	21	35
agefstm	Mother's age in years when first child was born	20.118	2.889	15	32
hispan	=1 if race is hispanic; = 0 if race is black	0.593	0.491	0	1
nonmomi	Non-mom's labor income	31.806	20.375	0	157.4
edu	Education = Number of schooling years	11.005	3.305	0	20
samesex	=1 if the 1st 2 kids have the same sex; = 0 otherwise	0.503	0.5	0	1
multi2nd	=1 if the 2nd birth is twin; =0 otherwise	0.009	0.093	0	1

Dependent Variable - Kidno	Linear model (OLS)	Negative Binomial II model (MLE)
edu	-0.065	-0.078
	(0.002)	(0.002)
age	0.096	0.119
	(0.002)	(0.002)
agefstm	-0.114	-0.156
	(0.002)	(0.003)
hispan	0.036	0.045
	(0.010)	(0.015)
nonmomi	-0.002	-0.003
	(0.000)	(0.000)
samesex	0.075	0.098
	(0.010)	(0.013)
multi2nd	0.786	0.728
	(0.052)	(0.045)
constant	0.911	0.013
	(0.042)	(0.067)

 Table A.16: First-stage Estimates using Instrumental Variables

Note: Figures in parentheses are robust standard errors.

Model	Linear	Tobit		Fractional Probit	
Estimation Method	OLS	MLE		QMLE	
	Coefficient	Coefficient	APE	Coefficient	APE
kidno (continuous)	-0.019	-0.034	-0.0225	-0.099	-0.0202
	(0.0007)	(0.0013)	(0.0008)	(0.0040)	(0.0008)
0-1			-0.0231		-0.0207
			(0.0008)		(0.0008)
1-2			-0.0207		-0.0185
			(0.0007)		(0.0007)
2-3			-0.0183		-0.0163
			(0.0005)		(0.0005)
edu	0.004	0.008	0.005	0.022	0.005
	(0.0002)	(0.0004)	(0.0002)	(0.0010)	(0.0002)
age	0.005	0.008	0.006	0.024	0.005
	(0.0002)	(0.0003)	(0.0002)	(0.0010)	(0.0002)
agefstm	-0.006	-0.01	-0.007	-0.03	-0.006
	(0.0003)	(0.0004)	(0.0003)	(0.0010)	(0.0003)
hispan	-0.032	-0.052	-0.034	-0.15	-0.031
	(0.0010)	(0.0022)	(0.0014)	(0.0070)	(0.0013)
nonmomi	-0.0003	-0.0006	-0.0004	-0.002	-0.0004
	(0.0000)	(0.0001)	(0.0000)	(0.0002)	(0.0000)

Table A.17: Estimates Assuming Number of Kids is Conditionally Exogenous

Note: Figures in parentheses under the Coefficient columns are robust standard errors. Figures in parentheses under the APE columns are bootstrapped standard errors.

Model	Linear	Tobit	(BS)	Fraction	al Probit	Fraction	al Probit	Fraction	al Probit
Estimation Method	2SLS	M	LE	QML	E-PW	QM	ILE	N	LS
				Kidno is as	sumed cont.				
	Coef.	Coef.	APE	Coef.	APE	Coef.	APE	Coef.	APE
kidno (continuous)	-0.016	-0.027	-0.018	-0.078	-0.016	-0.081	-0.017	-0.081	-0.017
	(0.0070)	(0.0130)	(0.0080)	(0.0370)	(0.0080)	(0.0070)	(0.0010)	(0.0070)	(0.0010)
0-1			-0.018		-0.016		-0.017		-0.017
			(0.0080)		(0.0080)		(0.0010)		(0.0010)
1-2			-0.017		-0.015		-0.015		-0.015
			(0.0070)		(0.0070)		(0.0010)		(0.0010)
2-3			-0.015		-0.014		-0.014		-0.014
			(0.0060)		(0.0050)		(0.0010)		(0.0010)
edu	0.004	0.009	0.006	0.024	0.005	0.024	0.005	0.024	0.005
	(0.0005)	(0.0009)	(0.0006)	(0.0020)	(0.0005)	(0.0010)	(0.0005)	(0.0010)	(0.0005)
age	0.005	0.008	0.005	0.022	0.004	0.021	0.004	0.021	0.004
	(0.0007)	(0.0010)	(0.0008)	(0.0040)	(0.0008)	(0.0010)	(0.0008)	(0.0010)	(0.0008)
agefstm	-0.006	-0.01	-0.006	-0.028	-0.006	-0.027	-0.005	-0.027	-0.005
	(0.0008)	(0.0010)	(0.0010)	(0.0040)	(0.0009)	(0.0020)	(0.0008)	(0.0020)	(0.0008)
hispan	-0.032	-0.052	-0.034	-0.15	-0.031	-0.151	-0.031	-0.151	-0.031
	(0.0010)	(0.0020)	(0.0010)	(0.0070)	(0.0010)	(0.0070)	(0.0010)	(0.0070)	(0.0010)
nonmomi	-0.0003	-0.0005	-0.0004	-0.002	-0.0003	-0.002	-0.0003	-0.002	-0.0003
	(0.00004)	(0.00006)	(0.00004)	(0.00002)	(0.00004)	(0.00020)	(0.00004)	(0.00020)	(0.00004)

Note: Figures in parentheses under the Coefficient columns are robust standard errors. Figures in parentheses under the APE columns are bootstrapped standard errors; those under the APEs for a count endogenous variable with the QMLE and NLS methods are computed standard errors.

Appendix B

TABLES AND FIGURES FOR CHAPTER 2

Variable Description	Mean	Standard deviation
Annual Hours	1105.7	886.52
Experience (years)	11.89	7.71
Education (years)	12.94	2.27
Age (years)	41.42	10.18
Number of children aged 0-2	0.13	0.37
Number of children aged 3-5	0.18	0.42
Number of children aged 6-17	0.84	1.01
Married (= 1 if married)	0.88	0.32
Husband's employment status (=1 if working)	0.82	0.39
Non-wife income (thousand dollars)	36,622.4	41,704
Number of observations	11,232	
Number of women	864	
Number of years	13	

 Table B.1: Summary Statistics

Dependent variable: Female Working Experience				
Number of children aged 0-2	-0.442**			
6	[0.200]			
Number of children aged 3-5	-0.707***			
6	[0.169]			
Number of children aged 6-17	-1.207***			
C C	[0.162]			
Years of schooling	0.470***			
-	[0.096]			
Married	-1.191			
	[1.186]			
Husband's work participation	-1.256			
	[0.913]			
Non-wife income	-0.00003***			
	[0.00001]			
Age	1.174***			
	[0.138]			
Age squared	-0.010***			
	[0.002]			
η	0.958			
Number of observations	11,232			
Number of women	864			
R-squared	0.37			
F-Statistics on IVs	196.26			

Table B.2: Determinants of Female Working Experience - First stage regressions

Note: *, **, ***: significant at 10%, 5% and 1% level respectively. Other explanatory variables include time dummies and time averages of all explanatory variables. Standard errors robust to heteroskedasticity and serial correlation are inside square brackets. Instrumental variables (IVs) are age and age squared.

Model	Dynamic Linear		Tobit
Estimation Method	GMM	Correlated	CRE with serial
		RE (CRE)	correlation correction
	[1]	[2]	[3]
Lagged Hours	0.857***	0.542***	0.492***
	[0.012]	[0.009]	[0.025]
Experience	4.683***	4.964**	13.207***
	[1.514]	[2.381]	[1.582]
Children 0-2	-37.657**	-73.537***	-148.978***
	[15.718]	[15.884]	[18.038]
Children 3-5	0.371	-44.292***	-97.080***
	[12.932]	[13.66]	[15.628]
Children 6-17	29.820***	46.139***	7.103
	[5.81]	[8.108]	[8.364]
Education	7.979**	-6.849	3.975
	[3.147]	[9.415]	[9.577]
Married	-134.671***	-253.253**	-234.204**
	[33.576]	[124.741]	[117.946]
Husband's work status	136.438***	205.205***	195.717***
	[26.787]	[25.982]	[28.254]
Non-wife income	-0.001***	0.001***	-0.001***
	[0.0004]	[0.0002]	[0.0002]
Initial Condition		0.161***	0.102***
		[0.038]	[0.011]
v2it		-59.413***	
		[3.643]	
v2it*			427.820***
			[47.673]
Observations	10368	10368	10368
Number of women	864	864	864

Table B.3: Estimating Dynamic Female Labor Supply, Second Stage Regressions, Experience is Treated as an Endogenous Variable

Note: *, **, ***: significant at 10%, 5% and 1% level respectively. \mathbf{z}_i and \mathbf{v}_{2i} are included in (2) and (3) but not reported in the table. The first stage residual in (3) is free of serial correlation. Standard errors corrected for the first stage estimation are inside square brackets.

Model	Dynamic Linear		Tobit
Estimation Method	GMM	CRE	CRE-SC
	[1]	[2]	[3]
Lagged Hours	0.857***	0.469***	0.434***
	[0.012]	[0.012]	[0.011]
Experience	4.683***	4.294	11.481***
	[1.514]	[3.074]	[2.662]
Children 0-2	-37.657**	-63.616***	-122.552***
	[15.718]	[17.612]	[14.848]
Children 3-5	0.371	-38.317**	-80.443***
	[12.932]	[12.932]	[10.343]
Children 6-17	29.820***	39.914***	6.262
	[5.81]	[10.10]	[8.722]
Education	7.979**	-5.925	3.504
	[3.147]	[12.38]	[9.669]
Married	-134.671***	-219.09	-251.11
	[33.576]	[278.695]	[244.441]
Husband's work status	136.438***	177.521***	161.88***
	[26.787]	[58.886]	[35.728]
Non-wife income	-0.001***	0.001***	-0.001***
	[0.0004]	[0.0003]	[0.0004]
Replications		100	100

Table B.4: Average Partial Effects on Female Labor Supply

Note: *, **, ***: significant at 10%, 5% and 1% level respectively. The figures inside square brackets are bootstrapped standard errors with 100 replications.



Figure B.1: Distribution of Women's Annual Hours of Work in 1980-1992



Figure B.2: Hours of Work vs. Experience



Figure B.3: Hours of Work vs. Number of Children 0-2



Figure B.4: Hours of Work vs. Number of Children 3-5


Figure B.5: Hours of Work vs. Number of Children 6-17

Appendix C

TABLES FOR CHAPTER 3

Variable Description	Mean	S.D.
Female labor participation	0.95	0.21
Annual Hours	1938.02	755.29
Education (years)	7.29	3.88
Age (years)	39.18	7.21
Newborn aged 0-2	0.1	0.3
Spouse's age (years)	41.9	7.48
Spouse's education (years)	8.2	3.86
Non-wife income (millions)	8.44	16.13
First child's gender	0.55	0.5
(=1 if a boy, =0 if a girl)		
First two children has same sex (=1 if yes, =0 if not)	0.56	0.5
Live in urban	0.18	0.38
Live with grandparent	0.09	0.29
Work on farm	0.84	0.37
Ethnic (=1 if major, =0 if minor)	0.79	0.4
Number of women	665	
Number of observation	1995	

Table C.1: Summary Statistics for the Whole Sample

Note: N=665 women. Years = 2004, 2006, 2008.

S.D. stands for standard deviation.

Variable Description	2004		2006		2008	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Annual Hours (hours)	1818.47	826.48	1842.79	839.42	1792.62	869.48
Female labor participation (=1 if work, =0 if not)	0.96	0.19	0.96	0.2	0.94	0.23
Newborn aged 0-1 (=1 if yes, =0 if not)	0.16	0.36	0.09	0.28	0.06	0.25
Education (years)	7.23	3.81	7.31	3.85	7.32	3.98
Age (years)	37.23	7.03	39.16	7	41.13	7.09
Non-wife income (million dongs)	6.32	14.77	7.89	13.05	11.1	19.53
Spouse's age (years)	39.97	7.32	41.9	7.32	43.83	7.3
Spouse's education (years)	8.12	3.81	8.18	3.8	8.3	3.96
First two children has same sex (=1 if yes, =0 if not)	0.59	0.49	0.57	0.5	0.52	0.5
Live in urban (=1 if yes, =0 if not)	0.17	0.38	0.18	0.38	0.19	0.39
Work on farm (=1 if yes, =0 if not)	0.86	0.35	0.83	0.38	0.82	0.38
Ethnic (=1 if major, =0 if minor)	0.8	0.4	0.8	0.4	0.79	0.41
Number of observations	66	5	66	5	66	5

Table C.2: Summary Statistics for Each Year in the Panel

	Fertility Equation	LFP Equation		
Dependent Variable	Newborn	LFP		
	[1]	[2]	[3]	
Explanatory Variable		Coefficient	APE/ATE	
Newborn	-	-0.136***	-0.127***	
		[0.05]	[0.03]	
Samesex	1.024***	-	-	
	[0.374]			
Non-wife income	0.008***	-0.001***	-0.001***	
	[0.003]	[0.0002]	[0.0002]	
Age	-0.302	-0.031*	-0.019*	
	[0.204]	[0.017]	[0.011]	
Age squared	0.002	0.0001*	0.0001	
	[0.002]	[0.00006]	[0.0001]	
Education	-0.06	-0.005	-0.003	
	[0.08]	[0.03]	[0.03]	
Husband's age	-	-0.007*	-0.005*	
		[0.004]	[0.003]	
Husband's age squared	-	0.0001	-0.0001	
		[0.0003]	[0.0002]	
Husband's education	-	-0.07*	-0.05*	
		[0.05]	[0.03]	
$Cov(v2it,v3it) = \rho$	-0.165			
	[0.04]			
Log likelihood	-866.48	3		
Number of observations	1995			

 Table C.3: Bivariate Probit Estimates of Fertility and LFP in the First Stage

Note: N=665, T=3. Time averages of explanatory variables and year dummies for 2006 and 2008 are included. Figures in square brackets are clustered standard errors to control for serial correlation across time.

*, **, ***: significant at 10%, 5% and 1% level respectively.

Explanatory Variable	Pooled	Pooled	Fixed	Fixed	SW	Proposed
	OLS	2SLS	Effect	Effect 2SLS	Procedure	Procedure
	[1]	[2]	[3]	[4]	[5]	[6]
Newborn	-0.134***	-0.232***	-0.164***	-0.277***	-0.308***	-0.345***
	[0.04]	[0.053]	[0.044]	[0.082]	[0.117]	[0.108]
Education	0.023***	0.024***	0.018*	0.014	0.019*	0.016**
	[0.004]	[0.006]	[0.01]	[0.011]	[0.012]	[0.008]
Age	-0.014	0.003	0.039	-0.002	0.035	-0.02**
	[0.017]	[0.046]	[0.033]	[0.067]	[0.037]	[0.01]
Age squared	0.0001	-0.0001	-0.001	0.0002	-0.0005	0.0001**
	[0.0002]	[0.0004]	[0.0004]	[0.001]	[0.0005]	[0.0001]
Non-wife income	0.001	0.001	0.001*	0.002*	0.002**	0.002***
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.0006]
Urban	0.090**	0.088^{**}	0.077	0.053	0.091**	0.095***
	[0.039]	[0.04]	[0.085]	[0.197]	[0.038]	[0.037]
Work on Farm	-0.318***	-0.317***	-0.101*	-0.104	-0.329***	-0.328***
	[0.037]	[0.037]	[0.054]	[0.066]	[0.033]	[0.034]
Ethnic	-0.143***	-0.138***	-0.176	-0.195	-0.142***	-0.143***
	[0.034]	[0.037]	[0.161]	[0.235]	[0.028]	[0.028]
R-square	0.1	0.08	0.07	0.03	0.1	0.11
Number of observations	1904	1904	1904	1904	1904	1904

Table C.4: Estimates for Log(Female Working Hours) Equation

Note: The dependent variable is log(hours), with 1904 observations of positive hours. Year dummy variables and time averages of explanatory variables are included. Standard errors are robust to serial correlation and heteroskedasticity. Standard errors in the SW and proposed procedure are corrected for the first-step estimation. *, **, ***: significant at 10%, 5% and 1% level respectively.

Appendix D

TECHNICALITIES FOR CHAPTER 1

D.1 Details of the QML Estimator

D.1.1 Asymptotic Variance for the Two-step Estimator

This section derives asymptotic standard errors for the QML estimator in the second step. The adjusted asymptotic standard errors for the NLS estimator can be derived in a similar way. In the first stage, we have: $y_2|z, a_1 \sim \text{Poisson}[\exp(z\delta_2 + a_1)]$ with the conditional density function:

$$f(y_2|\mathbf{z}, a_1) = \frac{[\exp(\mathbf{z}\delta_2 + a_1)]^{y_2 i} \exp\left[-\exp(\mathbf{z}\delta_2 + a_1)\right]}{y_2 !}.$$
 (D.1)

The unconditional density of y_2 conditioned only on z is obtained by integrating a_1 out of the joint density. That is:

$$f(y_2|\mathbf{z}) = \int_{a_1} f(y_2|\mathbf{z}, a_1) f(a_1) da_1, \tag{D.2}$$

in which $f(a_1) = \delta_0^{\delta_0} \exp(a_1)^{\delta_0 - 1} \exp(-\delta_0 \exp(a_1)) \Gamma^{-1}(\delta_0).$

Let $m = \exp(z\delta_2)$ and $c = \exp(a_1)$, then the conditional density is:

$$f(y_2|\mathbf{z}, a_1) = \frac{[mc]^{y_2} \exp[-mc]}{\Gamma(y_2 + 1)},$$

and the unconditional density is:

$$f(y_2|\mathbf{z}) = \int_0^\infty \frac{[mc]^{y_2} \exp\left[-mc\right]}{\Gamma(y_2+1)} \frac{\delta_0^{\delta_0} c^{\delta_0-1} \exp(-\delta_0 c))}{\Gamma(\delta_0)} dc.$$

This is equivalent to:

$$f(y_2|\mathbf{z}) = \frac{[m]^{y_2} \delta_0^{\delta_0}}{\Gamma(y_2+1)\Gamma(\delta_0)} \int_0^\infty \exp[-c(m+\delta_0)] c^{y_2 i + \delta_0 - 1} dc,$$

or

$$f(y_2|\mathbf{z}) = \frac{[m]^{y_2} \delta_0^{\delta_0}}{\Gamma(y_2+1)\Gamma(\delta_0)} \frac{\Gamma(y_2+\delta_0)}{(m+\delta_0)^{(y_2+\delta_0)}}$$

Define $h = \frac{\delta_0}{m + \delta_0}$ results in:

$$f(y_2|\mathbf{z}) = \frac{\Gamma(y_2 + \delta_0)h^{\delta_0}(1 - h)^{y_2}}{\Gamma(y_2 + 1)\Gamma(\delta_0)},$$
(D.3)

where $y_2 = 0, 1, ...$ and $\delta_0 > 0$, which is the density function for the negative binomial distribution.

The log-likelihood for observation *i* is:

$$l_i(\delta_2, \delta_0) = \delta_0 \ln\left[\frac{\delta_0}{\delta_0 + \exp(\mathbf{z}_i \delta_2)}\right] + y_{2i} \ln\left[\frac{\exp(\mathbf{z}_i \delta_2)}{\delta_0 + \exp(\mathbf{z}_i \delta_2)}\right] + \ln\left[\frac{\Gamma(y_{2i} + \delta_0)}{\Gamma(y_{2i} + 1)\Gamma(\delta_0)}\right].$$
(D.4)

For all observations:

$$L(\delta_2, \delta_0) = \sum_{i=1}^{N} l_i(\delta_2, \delta_0).$$
 (D.5)

We can estimate jointly δ_2 and δ_0 by maximum likelihood estimation method.

Let $\gamma = (\delta_2, \delta_0)'$ has the dimension of (L+1) where *L* is the dimension of δ_2 which is the sum of *K* and the number of instruments, under standard regularity conditions, we have:

$$\sqrt{N}(\hat{\gamma} - \gamma) = N^{-1/2} \sum_{i=1}^{N} \mathbf{r}_{i2} + o_p(1),$$
 (D.6)

where

$$\mathbf{r}_{i2} = \begin{bmatrix} -\mathbf{A}_{01}^{-1} \mathbf{s}_{01} \\ -\mathbf{A}_{02}^{-1} \mathbf{s}_{02} \end{bmatrix},$$
 (D.7)

in which
$$\mathbf{s}_0 = \begin{pmatrix} \nabla_{\delta_2} l_i \\ \nabla_{\delta_0} l_i \end{pmatrix} = \begin{pmatrix} \mathbf{s}_{01} \\ \mathbf{s}_{02} \end{pmatrix}$$
, and
 $\mathbf{A}_0 = E(\nabla_{\gamma}^2 l_i) = E\begin{pmatrix} \nabla_{\delta_2}^2 l_i \\ \nabla_{\delta_0}^2 l_i \end{pmatrix} = E\begin{pmatrix} \mathbf{H}_{01} \\ \mathbf{H}_{02} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{01} \\ \mathbf{A}_{02} \end{pmatrix}$.

After taking the first derivative and the second derivative, we have:

$$\mathbf{s}_{01} = \frac{\mathbf{z}_i' \delta_0(y_{2i} - \exp(\mathbf{z}_i \delta_2))}{\delta_0 + \exp(\mathbf{z}_i \delta_2)},\tag{D.8}$$

$$\mathbf{H}_{01} = -\frac{\mathbf{z}_i' \mathbf{z}_i \delta_0 \exp(\mathbf{z}_i \delta_2)}{\delta_0 + \exp(\mathbf{z}_i \delta_2)},\tag{D.9}$$

$$\mathbf{s}_{02} = \ln(\frac{\delta_0}{\delta_0 + \exp(\mathbf{z}_i \delta_2)}) + \frac{\exp(\mathbf{z}_i \delta_2) - y_{2i}}{\delta_0 + \exp(\mathbf{z}_i \delta_2)} + \frac{\Gamma'(y_{2i} + \delta_0)}{\Gamma(y_{2i} + \delta_0)} - \frac{\Gamma'(\delta_0)}{\Gamma(\delta_0)}, \tag{D.10}$$

$$\mathbf{H}_{02} = \mathbf{H}_{021} + \mathbf{H}_{022},\tag{D.11}$$

where

$$\mathbf{H}_{021} = \frac{\exp(\mathbf{z}_i \delta_2)}{\delta_0 \left[\delta_0 + \exp(\mathbf{z}_i \delta_2)\right]} - \frac{\exp(\mathbf{z}_i \delta_2) - y_{2i}}{\left[\delta_0 + \exp(\mathbf{z}_i \delta_2)\right]^2}$$

and

$$\mathbf{H}_{022} = \frac{\Gamma''(y_{2i} + \delta_0)\Gamma(y_{2i} + \delta_0) - \left[\Gamma'(y_{2i} + \delta_0)\right]^2}{\left[\Gamma(y_{2i} + \delta_0)\right]^2} - \frac{\Gamma''(\delta_0)\Gamma(\delta_0) - \left[\Gamma'(\delta_0)\right]^2}{\left[\Gamma(\delta_0)\right]^2}$$

where \mathbf{s}_{01} and \mathbf{H}_{01} are $L \times 1$ and $L \times L$ matrices; \mathbf{s}_{012} and \mathbf{H}_{02} are 1×1 and 1×1 matrices. $\mathbf{r}_{i2}(\gamma)$ has the dimension of $(L+1) \times 1$.

With the two-step M-estimator, the asymptotic variance of $\sqrt{N}(\hat{\theta} - \theta)$ must be adjusted to account for the first-stage estimation of $\sqrt{N}(\hat{\gamma} - \gamma)$ (see more in 12.4.2 of chapter 12, Wooldridge, 2002).

The score of the QML (or the gradient) for observation *i* with respect to θ is:

$$\begin{aligned} \mathbf{s}_{i}(\theta;\gamma) &= \nabla_{\theta}l_{i}(\theta), \\ &= y_{1i}\frac{\nabla_{\theta}\mu_{i}}{\mu_{i}} - (1-y_{1i})\frac{\nabla_{\theta}\mu_{i}}{1-\mu_{i}}, \\ &= \frac{y_{1i}\nabla_{\theta}\mu_{i}(1-\mu_{i}) - \mu_{i}(1-y_{1i})\nabla_{\theta}\mu_{i}}{\mu_{i}(1-\mu_{i})}, \\ &= \frac{y_{1i}\nabla_{\theta}\mu_{i} - \mu_{i}\nabla_{\theta}\mu_{i}}{\mu_{i}(1-\mu_{i})}, \\ &= \frac{(y_{1i}-\mu_{i})\nabla_{\theta}\mu_{i}}{\mu_{i}(1-\mu_{i})}, \\ &= \frac{(y_{1i}-\mu_{i})}{\mu_{i}(1-\mu_{i})}\int_{-\infty}^{+\infty}\frac{\partial\Phi(\mathbf{g}_{i}\theta)}{\partial\theta}f(a_{1}|y_{2},\mathbf{z})da_{1} \\ \mathbf{s}_{i}(\theta;\gamma) &= \frac{(y_{1i}-\mu_{i})}{\mu_{i}(1-\mu_{i})}\int_{-\infty}^{+\infty}\mathbf{g}_{i}'\phi(\mathbf{g}_{i}\theta)f(a_{1}|y_{2},\mathbf{z})da_{1}, \end{aligned}$$
(D.12)

where $\mathbf{g}_i = (y_{2i}, \mathbf{z}_{1i}, a_{1i})$ and $\boldsymbol{\theta} = (\alpha_1, \delta_1, \eta_1)'$ and $\boldsymbol{\theta}$ has the dimension of K + 2.

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{A}_1^{-1}(N^{-1/2}\sum_{i=1}^N \mathbf{r}_{i1}(\boldsymbol{\theta}; \boldsymbol{\gamma})) + o_p(1),$$
(D.13)

$$\mathbf{A}_{1} = E\left[-\bigtriangledown_{\boldsymbol{\theta}} \mathbf{s}_{i}(\boldsymbol{\theta};\boldsymbol{\gamma})\right],$$

$$= E\left[\frac{(\bigtriangledown_{\boldsymbol{\theta}}\boldsymbol{\mu}_{i})'\bigtriangledown_{\boldsymbol{\theta}}\boldsymbol{\mu}_{i}}{\boldsymbol{\mu}_{i}(1-\boldsymbol{\mu}_{i})}\right],$$

$$= E\left[\frac{1}{\boldsymbol{\mu}_{i}(1-\boldsymbol{\mu}_{i})}\mathbf{B}'\mathbf{B}\right].$$

$$\mathbf{\hat{A}}_{1} = N^{-1}\sum_{i=1}^{N}\left[\frac{1}{\boldsymbol{\mu}_{i}(1-\boldsymbol{\mu}_{i})}\mathbf{\hat{B}}'\mathbf{\hat{B}}\right],$$
(D.14)

where $\mathbf{B} = \int_{-\infty}^{+\infty} \mathbf{g}'_i \phi(\mathbf{g}_i \theta) f(a_1 | y_2, \mathbf{z}) da_1.$

$$\mathbf{r}_{i1}(\boldsymbol{\theta};\boldsymbol{\gamma}) = \mathbf{s}_{i}(\boldsymbol{\theta};\boldsymbol{\gamma}) - \mathbf{F}_{1}\mathbf{r}_{i2}(\boldsymbol{\gamma}),$$
$$\mathbf{\hat{r}}_{i1}(\boldsymbol{\theta};\boldsymbol{\gamma}) = \mathbf{\hat{s}}_{i}(\boldsymbol{\theta};\boldsymbol{\gamma}) - \mathbf{\hat{F}}_{1}\mathbf{\hat{r}}_{i2}(\boldsymbol{\gamma}),$$
(D.15)

where $\mathbf{r}_{i1}(\theta; \gamma)$, $\mathbf{s}_i(\theta; \gamma)$ are $(K+2) \times 1$ matrices, and $\mathbf{r}_{i2}(\gamma)$ and \mathbf{F}_1 are $(L+1) \times 1$ and $(K+2) \times (L+1)$ matrices, \mathbf{A}_1 is a $(K+2) \times (K+2)$ matrix.

$$\begin{split} \mathbf{F}_{1} &= E[\bigtriangledown_{\gamma} \mathbf{s}_{i}(\theta; \gamma)] = E\begin{pmatrix} \bigtriangledown_{\delta_{2}} \mathbf{s}_{i}(\theta; \gamma) \\ \bigtriangledown_{\delta_{0}} \mathbf{s}_{i}(\theta; \gamma) \end{pmatrix}, \\ E\left(\bigtriangledown_{\delta_{2}} \mathbf{s}_{i}(\theta; \gamma)\right) &= E\left[\frac{-1}{\mu_{i}(1-\mu_{i})} \mathbf{B}\left(\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_{i}\theta) \frac{\partial f(a_{1}|y_{2}, \mathbf{z})}{\partial \delta_{2}} da_{1}\right)\right], \\ E\left(\bigtriangledown_{\delta_{0}} \mathbf{s}_{i}(\theta; \gamma)\right) &= E\left[\frac{-1}{\mu_{i}(1-\mu_{i})} \mathbf{B}\left(\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_{i}\theta) \frac{\partial f(a_{1}|y_{2}, \mathbf{z})}{\partial \delta_{0}} da_{1}\right)\right], \end{split}$$

$$\mathbf{\hat{F}}_{1} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} [\mu_{i}(1-\mu_{i})]^{-1} \mathbf{\hat{B}} \left(\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_{i} \mathbf{\hat{\theta}}) [\partial f(a_{1}|y_{2}, \mathbf{z})/\partial \delta_{2}] da_{1} \right) \\ [\mu_{i}(1-\mu_{i})]^{-1} \mathbf{\hat{B}} \left(\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_{i} \mathbf{\hat{\theta}}) [\partial f(a_{1}|y_{2}, \mathbf{z})/\partial \delta_{0}] da_{1} \right) \end{bmatrix},$$
(D.16)

where

$$\frac{\partial f(a_1|y_{2i}, \mathbf{z}_i)}{\partial \delta_2} = \frac{\mathbf{z}_i' \mathbf{P} \mathbf{C}[\delta_0 + \exp(\mathbf{z}_i \delta_2)]^{(y_{2i} + \delta_0 - 1)}}{\Gamma(y_{2i} + \delta_0)}, \tag{D.17}$$

in which

$$\mathbf{P} = -\exp(\mathbf{z}_i\delta_2 + a_1) + a_1(y_{2i} + \delta_0) - \delta_0\exp(a_1),$$

and

$$\mathbf{C} = \{(y_{2i} + \delta_0) \exp(\mathbf{z}_i \delta_2) - \exp(\mathbf{z}_i \delta_2 + a_1) [\delta_0 + \exp(\mathbf{z}_i \delta_2)]\}.$$

$$\frac{\partial f(a_1|y_2, \mathbf{z})}{\partial \delta_0} = f(a_1|y_2, \mathbf{z})\mathbf{D},\tag{D.18}$$

in which $\mathbf{D} = a_1 - a_1 \exp(a_1) + \ln(\delta_0 + \exp(\mathbf{z}_i \delta_2)) + \frac{y_{2i} + \delta_0}{\delta_0 + \exp(\mathbf{z}_i \delta_2)} - \Gamma'(y_{2i} + \delta_0)$ and

$$f(a_1|y_2, \mathbf{z}) = \frac{\exp(\mathbf{P})[\delta_0 + \exp(\mathbf{z}\delta_2)]^{(y_2 + \delta_0)}}{\Gamma(y_2 + \delta_0)}$$

Therefore, we can obtain the asymptotic variance of the two-step estimator as:

$$Avar\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{A}_1^{-1} Var[\mathbf{r}_{i1}(\boldsymbol{\theta};\boldsymbol{\gamma})] \mathbf{A}_1^{-1}, \qquad (D.19)$$

and the estimator of this variance is:

$$\widehat{Avar}(\widehat{\theta}) = \frac{1}{N} \widehat{\mathbf{A}}_1^{-1} \left(N^{-1} \sum_{i=1}^N \widehat{\mathbf{r}}_{i1} \widehat{\mathbf{r}}_{i1}' \right) \widehat{\mathbf{A}}_1^{-1}.$$
(D.20)

The asymptotic standard errors are obtained by the square roots of the diagonal elements of this matrix.

D.1.2 Asymptotic Variance for the APEs

First, we need to obtain the asymptotic variance of $\sqrt{N}(\hat{\psi} - \psi)$ for continuous explanatory variable where:

$$\hat{\boldsymbol{\psi}} = \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}\hat{\boldsymbol{\theta}}) f(a_1 | y_2, \mathbf{z}; \hat{\boldsymbol{\theta}}) da_1 \right) \hat{\boldsymbol{\theta}}, \tag{D.21}$$

is the vector of scaled coefficients times the scaled factor in the APE section

$$\boldsymbol{\psi} = \left(\int_{-\infty}^{+\infty} \boldsymbol{\phi}(\mathbf{g}\boldsymbol{\theta}) f(a_1 | y_2, \mathbf{z}; \boldsymbol{\theta}) da_1 \right) \boldsymbol{\theta}, \tag{D.22}$$

is the vector of scaled population coefficients times the mean response.

If y_2 is treated as a continuous variable:

$$\widehat{APE} = \left(\int_{-\infty}^{+\infty} \phi(\hat{\alpha}_1 y_2 + \mathbf{z}_1 \hat{\delta}_1 + \hat{\eta}_1 a_1) f(a_1 | y_2, \mathbf{z}; \hat{\theta}) da_1 \right) \hat{\alpha}_1.$$

For a continuous variable z_{11} :

$$\widehat{APE} = \left(\int_{-\infty}^{+\infty} \phi(\hat{\alpha}_1 y_2 + \mathbf{z}_1 \hat{\delta}_1 + \hat{\eta}_1 a_1) f(a_1 | y_2, \mathbf{z}; \hat{\theta}) da_1 \right) \hat{\delta}_{11}.$$

Using problem 12.12 in Wooldridge (2002), and let $\hat{\pi} = (\hat{\theta}', \hat{\delta}'_2, \hat{\delta}'_0)'$ we have:

$$\sqrt{N}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) = N^{-1/2} \sum_{i=1}^{N} \left[\mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi) - \boldsymbol{\psi} \right] + E[\nabla_{\pi} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)] \sqrt{N}(\hat{\boldsymbol{\pi}} - \pi) + o_p(1), \qquad (D.23)$$

where

$$\mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi) = \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_i \theta) f(a_1 | y_2, \mathbf{z}; \theta) da_1 \right) \theta,$$

and

$$f(a_1|y_2,z) = f(a_1;\boldsymbol{\delta}_0,\boldsymbol{\delta}_0) \left(\frac{\boldsymbol{\delta}_0 + \exp(\mathbf{z}\boldsymbol{\delta}_2)}{\boldsymbol{\delta}_0 + \exp(\mathbf{z}\boldsymbol{\delta}_2 + a_1)}\right)^{\boldsymbol{\delta}_0 + y_2} [\exp(a_1)]^{y_2}.$$

First, we need to find $\sqrt{N}(\hat{\pi} - \pi)$

$$\sqrt{N}(\hat{\pi} - \pi) = N^{-1/2} \sum_{i=1}^{N} \begin{pmatrix} \mathbf{A}_{1}^{-1} \mathbf{r}_{i1} \\ \mathbf{r}_{i2} \end{pmatrix} + o_{p}(1),$$
$$\sqrt{N}(\hat{\pi} - \pi) = N^{-1/2} \sum_{i=1}^{N} \mathbf{k}_{i} + o_{p}(1).$$
(D.24)

Thus the asymptotic variance of $\sqrt{N}(\hat{\psi} - \psi)$ is:

$$Var\left\{\left[\left(\int_{-\infty}^{+\infty}\phi(\mathbf{g}_{i}\boldsymbol{\theta})f(a_{1}|y_{2i},\mathbf{z}_{i})da_{1}\right)\boldsymbol{\theta}-\boldsymbol{\psi}\right]+\mathbf{J}(\boldsymbol{\pi})\mathbf{k}_{i}\right\},\tag{D.25}$$

where $\mathbf{J}(\boldsymbol{\pi}) = E[\nabla_{\boldsymbol{\pi}} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \boldsymbol{\pi})].$

Next, we need to find $\nabla_{\theta} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)$; $\nabla_{\delta_2} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)$ and $\nabla_{\delta_0} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)$.

$$\nabla_{\boldsymbol{\theta}} \mathbf{j}(\mathbf{g}_{i}, \mathbf{z}_{i}, \pi) = \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_{i}\boldsymbol{\theta}) f(a_{1}|y_{2i}, \mathbf{z}_{i}) da_{1} \right) \mathbf{I}_{K+2} - \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_{i}\boldsymbol{\theta}) \left(\mathbf{g}_{i}\boldsymbol{\theta}\right) \left(\boldsymbol{\theta}\mathbf{g}_{i}\right) f(a_{1}|y_{2i}, \mathbf{z}_{i}) da_{1} \right),$$
(D.26)

where I_{K+2} is the identity matrix and (K+2) is the dimension of θ .

$$\nabla_{\delta_2} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi) = \theta \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_i \theta) \frac{\partial f(a_1 | y_{2i}, \mathbf{z}_i)}{\partial \delta_2} da_1 \right)', \tag{D.27}$$

where $\partial f(a_{1i}|y_{2i}, \mathbf{z}_i)/\partial \delta_2$ is defined in (D.17) and

$$\nabla_{\boldsymbol{\delta}_{0}}\mathbf{j}(\mathbf{g}_{i},\mathbf{z}_{i},\boldsymbol{\pi}) = \theta \left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_{i}\theta) \frac{\partial f(a_{1}|y_{2i},\mathbf{z}_{i})}{\partial \boldsymbol{\delta}_{0}} da_{1} \right)', \tag{D.28}$$

where $\partial f(a_{1i}|y_{2i}, \mathbf{z}_i)/\partial \delta_0$ is defined in (D.18). $\nabla_{\delta_2} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)$ is $(K+2) \times L$ matrix and $\nabla_{\delta_0} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi)$ is $(K+2) \times 1$ matrix.

Then,

$$\nabla_{\pi} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi) = \left[\nabla_{\theta} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi; \theta) | \nabla_{\delta_2} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi; \delta_2) | \nabla_{\delta_0} \mathbf{j}(\mathbf{g}_i, \mathbf{z}_i, \pi; \delta_0) \right], \quad (D.29)$$

and its expected value is estimated as:

$$\mathbf{\hat{J}} = \mathbf{J}(\mathbf{\hat{\pi}}) = N^{-1} \sum_{i=1}^{N} \left[\nabla_{\theta} \mathbf{j}(\mathbf{g}_{i}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\theta}}) | \nabla_{\delta_{2}} \mathbf{j}(\mathbf{g}_{i}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\delta}}_{2}) | \nabla_{\delta_{0}} \mathbf{j}(\mathbf{g}_{i}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\delta}}_{0}) \right].$$
(D.30)

Finally, $Avar\left[\sqrt{N}(\hat{\psi} - \psi)\right]$ is consistently estimated as:

$$\widehat{Avar}\left[\sqrt{N}(\widehat{\boldsymbol{\psi}}-\boldsymbol{\psi})\right] = N^{-1} \sum_{i=1}^{N} \left[\left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_{i}\widehat{\boldsymbol{\theta}}) f(a_{1}|y_{2i},\mathbf{z}_{i}) da_{1} \right) \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\psi}} + \widehat{\mathbf{J}}\widehat{\mathbf{k}}_{i} \right] \\ \times \left[\left(\int_{-\infty}^{+\infty} \phi(\mathbf{g}_{i}\widehat{\boldsymbol{\theta}}) f(a_{1}|y_{2i},\mathbf{z}_{i}) da_{1} \right) \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{\psi}} + \widehat{\mathbf{J}}\widehat{\mathbf{k}}_{i} \right]'.$$
(D.31)

where all quantities are evaluated at the estimators given above. The asymptotic standard error for any particular APE is obtained as the square root of the corresponding diagonal element of (D.31), divided by \sqrt{N} .

Now we obtain the asymptotic variance of $\sqrt{N}(\hat{\lambda} - \lambda)$ for a count endogenous variable where:

$$APE = E_{a_1}[\Phi(\alpha_1 y_2^{k+1} + z_1 \delta_1 + \eta_1 a_1) - \Phi(\alpha_1 y_2^k + z_1 \delta_1 + \eta_1 a_1)].$$
(D.32)

For example, $y_2^k = 0$ and $y_2^{k+1} = 1$.

$$\widehat{APE} = \left(\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_i^{k+1}\hat{\boldsymbol{\theta}}) f(a_1|y_2, \mathbf{z}; \hat{\boldsymbol{\theta}}) da_1 - \int_{-\infty}^{+\infty} \Phi(\mathbf{g}_i^k \hat{\boldsymbol{\theta}}) f(a_1|y_2, \mathbf{z}; \hat{\boldsymbol{\theta}}) da_1 \right),$$
(D.33)

$$\begin{split} Var\left[\sqrt{N}(\widehat{\lambda}-\lambda)\right] &= Var\sqrt{N}\left[(\widehat{\lambda}_{k+1}-\widehat{\lambda}_{k})-(\lambda_{k+1}-\lambda_{k})\right],\\ &= Var\sqrt{N}(\widehat{\lambda}_{k+1}-\lambda_{k+1})+Var\sqrt{N}(\widehat{\lambda}_{k}-\lambda_{k})\\ &-2Cov[\sqrt{N}(\widehat{\lambda}_{k+1}-\lambda_{k+1}),\sqrt{N}(\widehat{\lambda}_{k}-\lambda_{k})]. \end{split}$$

(1) We start with:

$$\sqrt{N}(\widehat{\lambda}_k - \lambda_k) = N^{-1/2} \sum_{i=1}^N \left(\mathbf{j}(\mathbf{g}_i^k, \mathbf{z}_i, \pi) - \lambda_k \right)$$

+
$$E[\nabla_{\pi}\mathbf{j}(\mathbf{g}_{i}^{k},\mathbf{z}_{i},\pi)]\sqrt{N}(\hat{\pi}-\pi)+o_{p}(1),$$
 (D.34)

where $\mathbf{j}(\mathbf{g}_i^k, \mathbf{z}_i, \pi) = \int_{-\infty}^{+\infty} \Phi(\mathbf{g}_i^k \theta) f(a_1 | y_{2i}, \mathbf{z}_i) da_1.$

$$\widehat{Var}\left[\sqrt{N}(\widehat{\lambda}_k - \lambda_k)\right] = N^{-1} \sum_{i=1}^{N} \left[\int_{-\infty}^{+\infty} \Phi(\mathbf{g}_i^k \widehat{\theta}) f(a_1 | y_{2i}, \mathbf{z}_i) da_1 - \widehat{\lambda}_k + \widehat{\mathbf{J}} \widehat{\mathbf{k}}_i \right]^2, \quad (D.35)$$

in which the notations of $\hat{\mathbf{k}}_i$ is the same as (D.24) and $\hat{\mathbf{J}}$ is defined as follows:

$$\mathbf{\hat{J}} = \mathbf{J}(\mathbf{\hat{\pi}}) = N^{-1} \sum_{i=1}^{N} \left[\nabla_{\theta} \mathbf{j}(\mathbf{g}_{i}^{k}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\theta}}) | \nabla_{\delta_{2}} \mathbf{j}(\mathbf{g}_{i}^{k}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\delta}}_{2}) | \nabla_{\delta_{0}} \mathbf{j}(\mathbf{g}_{i}^{k}, \mathbf{z}_{i}, \pi; \mathbf{\hat{\delta}}_{0}) \right],$$
(D.36)

$$\nabla_{\boldsymbol{\theta}} \mathbf{j}(\mathbf{g}_{i}^{k}, \mathbf{z}_{i}, \boldsymbol{\pi}; \boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \mathbf{g}_{i}^{k\prime} \phi(\mathbf{g}_{i}^{k} \widehat{\boldsymbol{\theta}}) f(a_{1} | y_{2i}, \mathbf{z}_{i}) da_{1},$$
(D.37)

$$\nabla_{\delta_{2}} \mathbf{j}(\mathbf{g}_{i}^{k}, \mathbf{z}_{i}, \pi; \delta_{2}) = \int_{-\infty}^{+\infty} \Phi(\mathbf{g}_{i}^{k} \widehat{\theta}) \frac{\partial f(a_{1} | y_{2i}, \mathbf{z}_{i})}{\partial \delta_{2}} da_{1}, \tag{D.38}$$

$$\nabla_{\delta_0} \mathbf{j}(\mathbf{g}_i^k, \mathbf{z}_i, \pi; \delta_0) = \int_{-\infty}^{+\infty} \Phi(\mathbf{g}_i^k \widehat{\theta}) \frac{\partial f(a_1 | y_{2i}, \mathbf{z}_i)}{\partial \delta_0} da_1.$$
(D.39)

(2) $\widehat{Var}\left[\sqrt{N}(\widehat{\lambda}_{k+1} - \lambda_{k+1})\right]$ is obtained in a similar way as (1).

(3) Using the formula: Cov(x,y) = E(xy) - ExEy and getting the estimator of this Covariance with the notice that $E(\widehat{\lambda}_k) = \lambda_k$, after some algebra, we have the estimator for this covariance is 0.

Adding (1), (2) and (3) together, we get:

$$\widehat{Var}\left[\sqrt{N}(\widehat{\lambda}-\lambda)\right] = \widehat{Var}\left[\sqrt{N}(\widehat{\lambda}_{k}-\lambda_{k})\right] + \widehat{Var}\left[\sqrt{N}(\widehat{\lambda}_{k+1}-\lambda_{k+1})\right].$$
(D.40)

The asymptotic standard error for APE of the count endogenous variable is obtained as the square root of the corresponding diagonal element of (D.40), divided by \sqrt{N}

D.2 Details of the Tobit Model's Estimators

This appendix shows how to obtain the average partial effects for Tobit models in both cases where y_2 is assumed exogenous and endogenous respectively.

Following the Smith-Blundell (1986) approach, the model with endogenous y_2 is written as:

$$y_1 = max(0, \alpha_1 y_2 + \mathbf{z}_1 \delta_1 + v_2 \xi_1 + e_1),$$
 (D.41)

where the reduced form of y_2 is:

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2, v_2 | \mathbf{z} \sim Normal(0, \boldsymbol{\Sigma}_2), \tag{D.42}$$

and $e_1|\mathbf{z}, v_2 \sim Normal(0, \sigma_e^2)$. The conditional mean of y_1 is:

$$E(y_1|\mathbf{z}, y_2, v_2) = \Phi[(\alpha_1 y_2 + \mathbf{z}_1 \delta_1 + v_2 \xi_1) / \sqrt{(1 + \sigma_e^2)}],$$

= $\Phi(\alpha_{1e} y_2 + \mathbf{z}_1 \delta_{1e} + v_2 \xi_{1e}).$

The Blundell-Smithprocedure for estimating $\alpha_1, \delta_1, \xi_1$ and σ_e^2 will then be:

- (i) Run the OLS regression of y_{i2} on z_i and save the residuals \hat{v}_{i2} , i = 1, 2, ..., N.
- (ii) Do Tobit of y_{i1} on y_{i2} , \mathbf{z}_{1i} and \hat{v}_{i2} to get $\hat{\alpha}_{1e}$, $\hat{\delta}_{1e}$, and $\hat{\xi}_{1e}$, i = 1, 2, ..., N.

APEs for Tobit model with exogenous or endogenous variable are obtained as follows:

* APE in Tobit Model with exogenous variable y_2

$$y_1 = \max(0, y_1^*), y_1^* = \alpha_1 y_2 + \mathbf{z}_1 \delta_1 + a_1, a_1 | y_2, \mathbf{z}_1 \sim N(0, \sigma^2).$$

The conditional mean is:

$$E(y_1|\mathbf{z}_1, y_2) = \Phi(\alpha_{1s}y_2 + \mathbf{z}_1\delta_{1s})(\alpha_1y_2 + \mathbf{z}_1\delta_1) + \sigma\phi(\alpha_{1s}y_2 + \mathbf{z}_1\delta_{1s}),$$
(D.43)

where $\alpha_{1s} = \frac{\alpha_1}{\sigma}, \delta_{1s} = \frac{\delta_1}{\sigma}$.

We define $E(y_1|\mathbf{z}_1, y_2) = m(y_2, \mathbf{z}_1, \theta_{1s}, \theta_1).$

For a continuous variable y_2 :

$$APE = \frac{\partial E(y_1|\mathbf{z}_1, y_2)}{\partial y_2} = \Phi(\alpha_{1s}y_2 + \mathbf{z}_1\delta_{1s})\alpha_1.$$
(D.44)

The estimator for this APE is:

$$\widehat{APE} = \frac{1}{N} \sum_{i=1}^{N} \Phi(\hat{\alpha}_{1s} y_{2i} + \mathbf{z}_{1i} \hat{\delta}_{1s}) \hat{\alpha}_1.$$
(D.45)

For a discrete variable y_2 with the two values c and c + 1:

$$APE = m(y_{2i} = c + 1) - m(y_{2i} = c),$$
(D.46)

and the estimator for this APE is:

$$\widehat{APE} = \frac{1}{N} \sum_{i=1}^{N} \hat{m}(y_{2i} = c+1) - \hat{m}(y_{2i} = c),$$
(D.47)

where $\hat{m}(y_{2i}=c) = \Phi(\hat{\alpha}_{1s}c + \mathbf{z}_{1i}\hat{\delta}_{1s})(\hat{\alpha}_1c + \mathbf{z}_{1i}\hat{\delta}_1) + \hat{\sigma}\phi(\hat{\alpha}_{1s}c + \mathbf{z}_{1i}\hat{\delta}_{1s}).$

* APE in Tobit Model with endogenous y_2 (Blundell-Smith 1986)

$$y_{1} = \max(0, y_{1}^{*}), y_{1}^{*} = \alpha_{1}y_{2} + \mathbf{z}_{1}\delta_{1} + \eta_{1}a_{1} + e_{1} = \alpha_{1}y_{2} + \mathbf{z}_{1}\delta_{1} + u_{1},$$

$$y_{2} = \mathbf{z}\delta_{2} + a_{1},$$

$$Var(a_{1}) = \sigma^{2}, e_{1}|z, a_{1} \sim N(0, \tau_{1}^{2}).$$

The standard method is to obtain APEs by computing the derivatives or the differences of:

$$E_{a_1}[m(\alpha_1 y_2 + \mathbf{z}_1 \delta_1 + \eta_1 a_1, \tau_1^2)], \qquad (D.48)$$

where $m(\alpha_1 y_2 + \mathbf{z}_1 \delta_1 + \eta_1 a_1, \tau_1^2) = m(\alpha_1 y_2 + \mathbf{z}_1 \delta_1, \eta_1^2 \sigma^2 + \tau_1^2).$

The conditional mean is:

$$E(y_1|\mathbf{z}_1, y_2) = \Phi(\alpha_{1s}y_2 + \mathbf{z}_1\delta_{1s})(\alpha_1y_2 + \mathbf{z}_1\delta_1) + \sqrt{\eta_1^2\sigma^2 + \tau_1^2}\phi(\alpha_{1s}y_2 + \mathbf{z}_1\delta_{1s}), \quad (D.49)$$

where $\alpha_{1s} = \frac{\alpha_1}{\sqrt{\eta_1^2 \sigma^2 + \tau_1^2}}, \delta_{1s} = \frac{\delta_1}{\sqrt{\eta_1^2 \sigma^2 + \tau_1^2}}.$ We define:

$$E(y_1|\mathbf{z}_1, y_2) = m(\alpha_1 y_2 + \mathbf{z}_1 \delta_1, \eta_1^2 \sigma^2 + \tau_1^2).$$
 (D.50)

Consistent estimators of APEs are resulted from the derivatives or the differences of $m(\hat{\alpha}_1 y_2 + \mathbf{z}_1 \hat{\delta}_1, \hat{\eta}_1^2 \hat{\sigma}^2 + \hat{\tau}_1^2)$ with respect to elements of (\mathbf{z}_1, y_2) where $\hat{\sigma}^2$ is the estimate of error variance from the first-stage OLS regression. \widehat{APE} with respect to z_1 :

$$\widehat{APE} = N^{-1} \sum_{i=1}^{N} \Phi(\hat{\alpha}_{1s} y_{2i} + \mathbf{z}_{1i} \hat{\delta}_{1s}) \hat{\alpha}_{1}, \qquad (D.51)$$

and \widehat{APE} with respect to y_2 :

$$\widehat{APE} = N^{-1} \sum_{i=1}^{N} \widehat{m}(y_{2i} = c+1) - \widehat{m}(y_{2i} = c),$$
(D.52)

where $\hat{m}(y_{2i} = c) = \Phi(\hat{\alpha}_{1s}c + \mathbf{z}_{1i}\hat{\delta}_{1s})(\hat{\alpha}_{1}c + \mathbf{z}_{1i}\hat{\delta}_{1}) + \sqrt{\hat{\eta}_{1}^{2}\hat{\sigma}^{2} + \hat{\tau}_{1}^{2}}\phi(\hat{\alpha}_{1s}c + \mathbf{z}_{1i}\hat{\delta}_{1s}).$

An alternative method is to get APEs by computing the derivatives or the differences of:

$$E_{a_1}[m(\alpha_1 y_2 + z_1 \delta_1 + \eta_1 a_1, \tau_1^2)], \tag{D.53}$$

where $m(z_1, y_2, a_1, \tau_1^2) = m(\mathbf{x}, \tau_1^2) = \Phi(\mathbf{x}/\tau_1)\mathbf{x} + \tau_1\phi(\mathbf{x}/\tau_1).$

 \widehat{APE} with respect to z_1 :

$$\widehat{APE} = N^{-1} \sum_{i=1}^{N} \Phi(\mathbf{x}/\sqrt{\widehat{\tau}_{1}^{2}}) \widehat{\delta}_{11}.$$
 (D.54)

 \widehat{APE} with respect to y_2 :

$$\widehat{APE} = N^{-1} \sum_{i=1}^{N} [\widehat{m}_{1} - \widehat{m}_{0}], \qquad (D.55)$$

where $\widehat{m}_0 = \widehat{m}[y_2 = 0]$ and $\mathbf{x} = \widehat{\alpha}_1 y_2 + z_1 \widehat{\delta}_1 + \widehat{\eta}_1 \widehat{a}_1$ and \widehat{a}_1 is the residual obtained from the first stage estimation.

For more details, see the Blundell-Smith procedure and the APEs in (Wooldridge, 2002, chapter 16).

D.3 Formula of the NLS estimation

In order to compare the NLS and the QML estimation, the basic framework is introduced as below.

The first stage is to estimate δ_2 and δ_0 by using the step-wise maximum likelihood of y_{i2} on \mathbf{z}_i in the Negative Binomial model. Obtain the estimated parameters $\hat{\delta}_2$ and $\hat{\delta}_0$. In the second stage, instead of using QMLE, we use the NLS of y_{i1} on y_{i2} , \mathbf{z}_{i1} to estimate α_1 , δ_1 and η_1 with the approximated conditional mean $\mu_i(\theta; y_2, \mathbf{z})$.

The NLS estimator of θ solves:

$$\begin{split} &\min_{\boldsymbol{\theta}\in\Theta} N^{-1} \sum_{i=1}^{N} \left[y_{1i} - \int_{-\infty}^{+\infty} \Phi(\alpha_1 y_{2i} + \mathbf{z}_{1i} \delta_1 + \eta_1 a_1) f(a_1 | y_2, \mathbf{z}) da_1 \right]^2, \\ & \text{or } \min_{\boldsymbol{\theta}\in\Theta} N^{-1} \sum_{i=1}^{N} [y_{1i} - \mu_i(\boldsymbol{\theta}; y_{2i}, \mathbf{z}_i)]^2 / 2. \end{split}$$

The score function can be written as:

$$\mathbf{s}_i = -(y_{1i} - \mu_i) \int_{-\infty}^{+\infty} \mathbf{g}'_i \phi(\mathbf{g}_i \boldsymbol{\theta}) f(a_1 | y_2, \mathbf{z}) da_1.$$
(D.56)

D.4 Derivation of the Heterogeneity Distribution

We are given $\exp(a_1)$ distributed as $Gamma(\delta_0, 1/\delta_0)$ using a single parameter δ_0 . We are interested in obtaining the density function of $Y = a_1$. Let $X = \exp(a_1)$. The density function of X is specified as follows:

$$f(X; \delta_0) = \frac{\delta_0^{\delta_0} X^{\delta_0 - 1} \exp(-\delta_0 X)}{\Gamma(\delta_0)}; \qquad X > 0, \ \delta_0 > 0.$$
(D.57)

Since X > 0 and $Y = \ln(X)$, $dX/dY = \exp(Y)$ and $Y \in (-\infty, \infty)$. The density function of *Y* will be derived as:

$$f(Y;\delta_0) = f[h(Y)] \left| \frac{dX}{dY} \right|; \qquad Y \in (-\infty,\infty),$$
(D.58)

where $f[h(Y)] = \frac{\delta_0^{\delta_0} \exp(Y)^{\delta_0 - 1} \exp[-\delta_0 \exp(Y)]}{\Gamma(\delta_0)}$.

Plug in $Y = a_1$, we get:

$$f(Y;\delta_0) = \frac{\delta_0^{\delta_0} \exp(a_1)^{\delta_0} \exp[-\delta_0 \exp(a_1)]}{\Gamma(\delta_0)},$$
(D.59)

which is equation (1.4).

Appendix E

TECHNICALITIES FOR CHAPTER 2

E.1 Asymptotic Variance of the Two-step Estimator

If the null hypothesis of no endogeneity and no serial correlation in the first stage is rejected, the standard errors in the second stage should be adjusted for the first stage estimation by using delta method or bootstrapping. In addition, we also need to get asymptotic standard errors for the average partial effects.

We start with the linear reduced form in the first stage:

$$y_{2it} = \mathbf{w}_{2it}^* \gamma_2 + v_{2it}^*, \tag{E.1}$$

where $\mathbf{w}_{2it}^* = (\mathbf{z}_{it}, \overline{\mathbf{z}}_i)$ is $1 \times (2L)$ vector of exogenous variables. Under standard regularity conditions, we have:

$$\sqrt{N}(\hat{\gamma}_2 - \gamma_2) = N^{-1/2} \sum_{i=1}^N \pi_{i2}(\gamma_2) + o_p(1), \qquad (E.2)$$

where

$$\pi_{i2} = \mathbf{A}_2^{-1} \mathbf{B}'_{2i} \mathbf{v}_{2i}^*, \tag{E.3}$$

and \mathbf{B}_{2i} is the $T \times (2L)$ matrix with *t*th row \mathbf{w}_{2it}^* , $\mathbf{A}_2 = E(\mathbf{B}'_{2i}\mathbf{B}_{2i})$ and \mathbf{v}_{2i}^* is a $T \times 1$ vector of reduced form errors.

Now we can write:

$$\mathbf{M} = E(y_{1it} | \mathbf{z}_i, y_{1i,t-1}, y_{2it}, y_{1i0}, \mathbf{w}_{2i}^*),$$

$$\mathbf{M} = m[\rho y_{1i,t-1} + \alpha y_{2it} + \mathbf{x}_{it}\beta + \theta_2 y_{1i0} + \mathbf{z}_i\theta_3 + (y_{2i}^* - \mathbf{w}_{2i}^*\gamma_2)\theta_4 + \theta_1(y_{2it}^* - \mathbf{w}_{2it}^*\gamma_2), \sigma_{s*}^2],$$

$$\mathbf{M} = m[\alpha y_{2it} + \mathbf{w}_{3it}\lambda_3 + \theta_1(y_{2it}^* - \mathbf{w}_{2it}^*\gamma_2) + (y_{2i}^* - \mathbf{w}_{2i}^*\gamma_2)\theta_4, \sigma_{s*}^2],$$
(E.4)

where $\mathbf{w}_{3it} = (y_{1i,t-1}, \mathbf{x}_{it}, y_{i0}, \mathbf{z}_i); \ \sigma_{s*}^2 = \sigma_{a_1}^{*2} + \sigma_{e_1}^{*2} \text{ and } \lambda_3 = (\rho, \beta', \theta_2, \theta_3')'.$

We collect all the parameters in **M** except for γ_2 into the parameter vector λ^* and abuse the notation that $\mathbf{w}_{it} = (y_{2it}, \mathbf{w}_{3it}, v_{2it}^*, \mathbf{v}_{2i}^*)$ in this part. In the previous part we use \mathbf{w}_{it}^* .

With the maximum likelihood in the second stage, the log likelihood for observation *i* in period time *t* is:

$$\mathbf{l}_{it}(\lambda^*; \sigma_{s*}^2) = \mathbf{1}[y_{1it} = 0] \log[1 - \Phi(\mathbf{w}_{it}\lambda^*/\sigma_{s*})] -\mathbf{1}[y_{1it} > 0] \{\log\phi[(y_{1it} - \mathbf{w}_{it}\lambda^*)/\sigma_{s*}] - \log(\sigma_{s*}^2)/2\}.$$
(E.5)

Using the notation: $\Phi(\mathbf{w}_i \lambda^* / \sigma_{s*}) = \Phi_i$; $\phi(\mathbf{w}_i \lambda^* / \sigma_{s*}) = \phi_i$ and the constant does not affect the maximization, we can rewrite this log likelihood as:

$$\mathbf{l}_{i}(\lambda^{*};\sigma_{s*}) = \mathbf{1}[y_{1i}=0]\log(1-\Phi_{i}) - \mathbf{1}[y_{1i}>0]\{\frac{1}{2}(y_{1i}-\mathbf{w}_{i}\lambda^{*})^{2}/\sigma_{s*}^{2} + \frac{1}{2}\log(\sigma_{s*}^{2})\}, \quad (E.6)$$

and we have the score as:

$$\mathbf{s}_{i}(\lambda^{*};\gamma_{2}) = \begin{bmatrix} \mathbf{s}_{i1} \\ \mathbf{s}_{i2} \end{bmatrix} = \begin{bmatrix} \nabla_{\lambda^{*}}\mathbf{l}_{i} \\ \nabla_{\sigma_{s^{*}}}\mathbf{l}_{i} \end{bmatrix}, \qquad (E.7)$$

and $\mathbf{s}_{i1} = -1[y_{1i} = 0](\phi_i \mathbf{w}_i) / \sigma_{s*}(1 - \Phi_i) + 1[y_{1i} > 0](y_{1i} - \mathbf{w}_i \lambda^*) \mathbf{w}_i / \sigma_{s*}^2$,

$$\mathbf{s}_{i2} = \mathbf{1}[y_{1i} = 0](\phi_i \mathbf{w}_i \lambda^*) / [2\sigma_{s*}^2(1 - \Phi_i)] + \mathbf{1}[y_{1i} > 0] \left[(y_{1i} - \mathbf{w}_i \lambda^*) / (2\sigma_{s*}^4) - 1 / (2\sigma_{s*}^2) \right].$$

With the two-step M-estimator, the asymptotic variance of $\sqrt{N}(\hat{\lambda}^* - \lambda^*)$ must be adjusted to account for the first-stage estimation of $\sqrt{N}(\hat{\gamma}_2 - \gamma_2)$ (see more in 12.4.2 of Chapter 12, Wooldridge, 2002). We can write:

$$\sqrt{N}(\widehat{\lambda}^* - \lambda^*) = \mathbf{A}_1^{-1} \left[N^{-1/2} \sum_{i=1}^N \pi_{i1}(\lambda^*; \gamma_2) \right] + o_p(1),$$
(E.8)

where

$$\mathbf{A}_1 = E[-\nabla_{\boldsymbol{\lambda}^*} \mathbf{s}_{i1}(\boldsymbol{\lambda}^*; \boldsymbol{\gamma}_2)], \tag{E.9}$$

and

$$\nabla_{\lambda^*} \mathbf{s}_{i1}(\lambda^*; \gamma_2) = -\sigma^{-2} \left\{ 1[y_{1i} = 0] \left([\phi_i^2 - \phi_i(1 - \Phi_i)\lambda^*] / (1 - \Phi_i)^2 \right) + 1[y_{1i} > 0] \right\} \mathbf{w}_i' \mathbf{w}_i,$$

and

$$\pi_{i1}(\lambda^*;\gamma_2) = \mathbf{s}_{i1}(\lambda^*;\gamma_2) - \mathbf{F}_1 \pi_{i2}(\gamma_2),$$
(E.10)

where

$$\mathbf{F}_{1} = E[\nabla_{\gamma_{2}}\mathbf{s}_{i1}(\lambda^{*};\gamma_{2})], \qquad (E.11)$$

or
$$\mathbf{F}_{1} = -E\left\{\sum_{t=1}^{T}[(1-\sigma_{s*})\phi_{it}\mathbf{w}_{it}\lambda^{*} + \Phi_{it}](\theta_{1}\mathbf{w}_{2it}^{*} + \theta_{4}\mathbf{w}_{2i}^{*})\right\}.$$

Therefore, we get:

$$Avar\left[\sqrt{N}(\widehat{\lambda}^* - \lambda^*)\right] = \mathbf{A}_1^{-1} \mathbf{V} \mathbf{A}_1^{-1}, \qquad (E.12)$$

where $\mathbf{V} = Var[\pi_{i1}(\lambda^*; \gamma_2)].$

A valid estimator of $Avar\left[\sqrt{N}(\widehat{\lambda}^* - \lambda^*)\right]$ is:

$$\widehat{\mathbf{A}}_{11} = \widehat{\mathbf{A}}_1^{-1} \left(N^{-1} \sum_{i=1}^N \widehat{\pi}_{i1} \widehat{\pi}'_{i1} \right) \widehat{\mathbf{A}}_1^{-1}, \tag{E.13}$$

where

$$\widehat{\mathbf{A}}_{1} = N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\sigma}^{-2} \left\{ 1[y_{1it} = 0] \left([\widehat{\phi}_{it}^{2} - \widehat{\phi}_{it}(1 - \widehat{\Phi}_{it})\widehat{\lambda}^{*}] / (1 - \widehat{\Phi}_{it})^{2} \right) + 1[y_{1it} > 0] \right\} \widehat{\mathbf{w}}_{it}' \widehat{\mathbf{w}}_{it},$$

and $\hat{\pi}_{i1} = \hat{\mathbf{s}}_{i1} - \hat{\mathbf{F}}_1 \hat{\pi}_{i2}$ in which $\hat{\pi}_{i2} = \hat{\mathbf{A}}_2^{-1} \hat{\mathbf{B}}_{2i}' \hat{\mathbf{v}}_{2i}^*$ and

$$\widehat{\mathbf{F}}_1 = -N^{-1} \sum_{i=1}^N \sum_{t=1}^T [(1 - \widehat{\sigma}_{s*}) \widehat{\phi}_{it} \widehat{\mathbf{w}}_{it} \widehat{\lambda}^* + \widehat{\Phi}_{it}] (\widehat{\theta}_1 \mathbf{w}_{2it}^* + \widehat{\theta}_4 \mathbf{w}_{2i}^*),$$

and the asymptotic variance of $\hat{\lambda}^*$ is:

$$\widehat{Avar(\widehat{\lambda}^*)} = \widehat{\mathbf{A}}_1^{-1} \widehat{\mathbf{Q}} \widehat{\mathbf{A}}_1^{-1} / N = \widehat{\mathbf{A}}_{11}, \qquad (E.14)$$

where $\widehat{\mathbf{Q}} = N^{-1} \sum_{i=1}^{N} \widehat{\pi}_{i1} \widehat{\pi}'_{i1}$.

We can derive $Avar\left[\sqrt{N}(\widehat{\sigma_{s*}^2} - \sigma_{s*}^2)\right]$ as the above procedure for the derivation of $Avar\left[\sqrt{N}(\widehat{\lambda}^* - \lambda^*)\right]$ and get \widehat{A}_{22} .

And denote, $\Psi \equiv (\lambda^*, \sigma_{s*}^2)'$, we can derive $\widehat{Avar} \left[\sqrt{N} (\widehat{\Psi} - \Psi) \right]$ as:

$$\widehat{Avar}\left[\sqrt{N}(\widehat{\Psi} - \Psi)\right] = \begin{bmatrix} \widehat{\mathbf{A}}_{11} & \widehat{\mathbf{A}}_{12} \\ \widehat{\mathbf{A}}_{21} & \widehat{\mathbf{A}}_{22} \end{bmatrix}, \quad (E.15)$$

where $\widehat{\mathbf{A}}_{22} = \widehat{\mathbf{A}}_2^{-1} \widehat{\mathbf{Q}} \widehat{\mathbf{A}}_2^{-1} / N$ and

$$\mathbf{A}_{2} = -\sigma^{-4} \{ (\mathbf{w}_{i}\lambda^{*}/\sigma_{s*})^{3} \phi_{i} + (\mathbf{w}_{i}\lambda^{*}/\sigma_{s*}) \phi_{i} - [(\mathbf{w}_{i}\lambda^{*}/\sigma_{s*})\phi_{i}^{2}/(1-\Phi_{i})] - 2\Phi_{i} \} / 4$$

and $\widehat{\mathbf{A}}_{12} = \widehat{\mathbf{A}}_{12}^{-1} \widehat{\mathbf{Q}} \widehat{\mathbf{A}}_{12}^{-1} / N$ and

$$\mathbf{A}_{12} = \sigma^{-3} \{ (\mathbf{w}_{\mathbf{i}} \lambda^* / \sigma_{\mathbf{s}*})^2 \phi_i + \phi_i - [(\mathbf{w}_{\mathbf{i}} \lambda^* / \sigma_{\mathbf{s}*}) \phi_i^2 / (1 - \Phi_i)] \} \mathbf{w}_{\mathbf{i}}' / 2.$$

Asymptotic Variance of the Average Partial Effects **E.2**

Next, we obtain the standard errors for the average partial effects as in equations (2.21) and (2.22).

$$\widehat{\varphi} = \widehat{\lambda}^* \left[(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \Phi(\mathbf{w}_{it} \widehat{\lambda}^* / \widehat{\sigma}_{s*}) \right], \qquad (E.16)$$

where $\widehat{\mathbf{w}}_{it} = (y_{2it}, \mathbf{w}_{3it}, \widehat{v}_{2it}^*, \widehat{\mathbf{v}}_{2i}^*)$.

$$\varphi = \lambda^* \left\{ T^{-1} \sum_{t=1}^{T} E\left[\Phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*}) \right] \right\}.$$
(E.17)

Then we need to compute the asymptotic variance of $\sqrt{N}(\widehat{\varphi} - \varphi)$.

Let $\widehat{\mu} = (\widehat{\lambda}^*; \widehat{\gamma}_2)'$ and $\mathbf{p}(\mathbf{w}_{it}, \mathbf{w}_{2it}^{*}, \mu) \equiv (T^{-1} \sum_{t=1}^{T} \left[\Phi(\left[\alpha y_{2it} + \mathbf{w}_{3it} \lambda_{3} + \theta_{1}(y_{2it}^{*} - \mathbf{w}_{2it}^{*} \gamma_{2}) + (y_{2i}^{*} - \mathbf{w}_{2i}^{*} \gamma_{2}) \theta_{4} \right] / \sigma_{s*}) \right]) \lambda^{*},$

we have:

$$\sqrt{N}(\widehat{\varphi} - \varphi) = N^{-1} \sum_{i=1}^{N} \left\{ \lambda^* (T^{-1} \sum_{t=1}^{T} [\Phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*})]) - \varphi \right\} + E[\nabla_{\mu} \mathbf{p}] \sqrt{N}(\widehat{\mu} - \mu) + o_p(1).$$
(E.18)

In which:

$$\sqrt{N}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) = N^{-1} \sum_{i=1}^{N} \mathbf{D}_i + o_p(1), \qquad (E.19)$$

where $\mathbf{D}_{i} = \begin{vmatrix} \mathbf{A}_{1}^{-1} \pi_{i1} \\ \pi_{i2} \end{vmatrix}$ and all matrix definitions were introduced in step 1 and a valid estimator

of \mathbf{D}_i is:

$$\widehat{\mathbf{D}} = \begin{bmatrix} \widehat{\mathbf{A}}_1^{-1} \widehat{\pi}_{i1} \\ \widehat{\pi}_{i2} \end{bmatrix}.$$
(E.20)

Therefore, the asymptotic variance of $\sqrt{N}(\widehat{\varphi} - \varphi)$ is:

$$\Omega = Var\left\{ \left[\lambda^* (T^{-1} \sum_{t=1}^T \left[\Phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*}) \right]) - \varphi \right] + \mathbf{PD} \right\},$$
(E.21)

in which $\Omega = Var(\mathbf{K} + \mathbf{PD})$ where $\mathbf{K} = \lambda^* (T^{-1} \sum_{t=1}^T [\Phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*})]) - \varphi$.

Hence, we can get

$$\widehat{\mathbf{K}} = \widehat{\lambda}^* (T^{-1} \sum_{t=1}^T \left[\Phi(\widehat{\mathbf{w}}_{it} \widehat{\lambda}^* / \widehat{\sigma}_{s*}) \right]) - \widehat{\varphi}.$$
(E.22)

The last job is to find the Jacobian **P** where $\mathbf{P} = E[\nabla_{\mu}\mathbf{p}]$

 $\mathbf{P} = [\mathbf{P}_1 | \mathbf{P}_2]$

$$\mathbf{P}_1 = \nabla_{\lambda^*} \mathbf{p} = T^{-1} \sum_{t=1}^T \left[\phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*}) \right] \left(\mathbf{w}_{it} \lambda^* / \sigma_{s*} \right) + \Phi(\mathbf{w}_{it} \lambda^* / \sigma_{s*}) \right],$$

and

$$\widehat{\mathbf{P}}_1 = T^{-1} \sum_{t=1}^{T} \left[\phi(\widehat{\mathbf{w}}_{it} \widehat{\lambda}^* / \widehat{\sigma}_{s*}) \right] (\widehat{\mathbf{w}}_{it} \widehat{\lambda}^* / \widehat{\sigma}_{s*}) + \Phi(\widehat{\mathbf{w}}_{it} \widehat{\lambda}^* / \widehat{\sigma}_{s*})],$$

or in short:

$$\widehat{\mathbf{P}}_1 = T^{-1} \sum_{t=1}^{T} \left[\phi(\widehat{\omega}) \widehat{\omega} + \Phi(\widehat{\omega}) \right],$$
(E.23)

where $\omega = \mathbf{w}_{it} \lambda^* / \sigma_{s*}$

$$\mathbf{P}_{2} = \nabla_{\gamma_{2}} \mathbf{p} = T^{-1} \sum_{t=1}^{T} \left[\phi(\mathbf{w}_{it} \lambda^{*} / \sigma_{s*}) \right] \left(-\theta_{1} \mathbf{w}_{2it}^{*} - \mathbf{w}_{2i}^{*} \theta_{4} \right) \lambda^{*} / \sigma_{s*} \right] \text{ and}$$
$$\widehat{\mathbf{P}}_{2} = T^{-1} \sum_{t=1}^{T} \left[\phi(\widehat{\omega}) \right] \left(-\widehat{\theta}_{1} \mathbf{w}_{2it}^{*} - \mathbf{w}_{2i}^{*} \widehat{\theta}_{4} \right) \widehat{\lambda}^{*} / \widehat{\sigma}_{s*} \right].$$
(E.24)

Therefore

$$\widehat{\mathbf{P}} = (NT)^{-1} \sum_{i=1}^{N} \left[T^{-1} \sum_{t=1}^{T} \left[\phi(\widehat{\omega}) \widehat{\omega} + \Phi(\widehat{\omega}) \right] | T^{-1} \sum_{t=1}^{T} \left[\phi(\widehat{\omega}) \right] (-\widehat{\theta}_{1} \mathbf{w}_{2it}^{*} - \mathbf{w}_{2i}^{*} \widehat{\theta}_{4}) \widehat{\lambda}^{*} / \widehat{\sigma}_{s*}) \right].$$
(E.25)

Finally, $Avar\left[\sqrt{N}(\widehat{\varphi} - \varphi)\right]$ is consistently estimated as:

$$\widehat{\boldsymbol{\Omega}} = N^{-1} \sum_{i=1}^{N} (\widehat{\mathbf{K}} - \widehat{\mathbf{P}} \widehat{\mathbf{D}}) (\widehat{\mathbf{K}} - \widehat{\mathbf{P}} \widehat{\mathbf{D}})'.$$
(E.26)

The asymptotic standard error for any particular APE is obtained as the square root of the corresponding diagonal element in the above expression, divided by \sqrt{N} .

Appendix F

TECHNICALITIES FOR CHAPTER 3

Derivation of Maximum likelihood estimator in the firs stage and the Asymptotic Variance in the second stage

F.1 Bivariate Probit Model in the First Stage

In the first stage, we estimate equation (3.11) and equation (3.12) simultaneously and get the log likelihood as in equation (3.23). Note that the model is qualitatively different from the usual bivariate probit model. In a simultaneous equations model (3.11-3.12), the second dependent variable y_{3it} appears on the right of the equation with the dependent variable y_{2it} . One can derive the following conditional mean and obtain the corresponding marginal effects of interest:

$$E(y_{2it}|W_i) = \Pr[y_{3it} = 1|W_i]E[y_{2it}|y_{3it} = 1, W_i] + \Pr[y_{3it} = 0|W_i]E[y_{2it}|y_{3it} = 0, W_i],$$
(F.1)

where

$$E[y_{2it}|y_{3it} = 1, W_i] = \Pr[y_{2it} = 1|y_{3it} = 1, W_i],$$
(F.2)

and

$$E[y_{2it}|y_{3it} = 0, W_i] = \Pr[y_{2it} = 1|y_{3it} = 0, W_i],$$
(F.3)

Therefore

$$E(y_{2it}|W_i) = \Phi_2(W_{3it}\gamma_3, W_{2it}\gamma_2 + \alpha_2; \rho) + \Phi_2(-W_{3it}\gamma_3, W_{2it}\gamma_2; -\rho).$$
(F.4)

To obtain the derivatives and Hessian, let us rewrite the log likelihood in a convenient way with $q_{2i} = 2y_{2i} - 1$ and $q_{3i} = 2y_{3i} - 1$ (which results in $q_{im} = 1$ if $y_{mi} = 1$ and $q_{im} = -1$ if $y_{mi} = 0$, for m = 2, 3):

$$\ln L_{it} = \ln \Phi_2(k_{i2}, k_{i3}; \pi), \tag{F.5}$$

where $k_{im} = q_{im}W_{mit}\gamma_m$ for m = 2, 3 (here the notation is abused under the note that $\gamma_2 = (\gamma'_2, \alpha_2)'$ and $\pi = q_{2i}q_{3i}\rho$.

The score function and the information matrix resulting from equation (F.5) are derived as follows:

$$s_{it}(\theta_1) = \nabla_{\theta_1} \ln L_{it}(\theta_1) = \begin{bmatrix} \frac{\partial \ln L_{it}(\theta_1)}{\partial \gamma_2} \\ \frac{\partial \ln L_{it}(\theta_1)}{\partial \gamma_2} \\ \frac{\partial \ln L_{it}(\theta_1)}{\partial \gamma_2} \\ \frac{\partial \ln L_{it}(\theta_1)}{\partial \gamma_2} \end{bmatrix}, \quad (F.6)$$

and

$$I(\theta_1) = -E\left[\nabla_{\theta_1}^2 \ln L_{it}(\theta_1)\right].$$
(F.7)

We have:

$$\partial \ln L_{it}(\theta_1) / \partial \gamma_3 = \Phi_2^{-1}(k_{i2}, k_{i3}; \pi)(q_{i3}W_{3it})g_{i3},$$
(F.8)

where $g_{i3} = \phi(k_{i3})\Phi\left[(k_{i2} - \pi k_{i3})(1 - \pi^2)^{-1/2}\right].$

$$\partial \ln L_{it}(\theta_1) / \partial \gamma_2 = \Phi_2^{-1}(k_{i2}, k_{i3}; \pi)(q_{i2}W_{2it})g_{i2}, \tag{F.9}$$

where $g_{i2} = \phi(k_{i2})\Phi\left[(k_{i3} - \pi k_{i2})(1 - \pi^2)^{-1/2}\right].$

$$\partial \ln L_{it}(\theta_1) / \partial \alpha_2 = \Phi_2^{-1}(k_{i2}, k_{i3}; \pi) q_{i2} g_{i2}, \tag{F.10}$$

and

$$\partial \ln L_{it}(\theta_1) / \partial \rho = \Phi_2^{-1}(k_{i2}, k_{i3}; \pi) q_{i2} q_{i3} \phi_2(k_{i2}, k_{i3}; \pi).$$
(F.11)

Therefore, the asymptotic variance of θ_1 is:

$$Avar(\theta_1) = C^{-1}VC^{-1}/N,$$
 (F.12)

where

$$C = N^{-1} \sum_{i=1}^{N} I(\theta_1),$$
(F.13)

and

$$V = N^{-1} \sum_{i=1}^{N} s_{it}(\theta_1) s_{it}(\theta_1)'.$$
 (F.14)

As a result, the estimator of the asymptotic variance of θ_1 is:

$$\widehat{Avar(\hat{\theta}_1)} = \widehat{C}^{-1}\widehat{V}\widehat{C}^{-1}/N, \tag{F.15}$$

and

$$\sqrt{N}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} Normal(\mathbf{0}, C^{-1}VC^{-1}), \tag{F.16}$$

or

$$\sqrt{N}(\widehat{\theta}_1 - \theta_1) = N^{-1/2} \sum_{i=1}^N \mathbf{r}_i(\theta_1) + o_p(1), \tag{F.17}$$

where

$$\mathbf{r}_i(\boldsymbol{\theta}_1) = -I(\boldsymbol{\theta}_1)^{-1} s_i(\boldsymbol{\theta}_1), \tag{F.18}$$

and

$$\hat{\mathbf{r}}_i(\theta_1) \equiv -I(\hat{\theta}_1)^{-1} s_i(\hat{\theta}_1). \tag{F.19}$$

F.2 Asymptotic Variance of the Two-step Estimator

The asymptotic variance of the second-stage parameters, θ_2 , needs to be corrected for general heterokedasticity, serial correlation and first-stage estimation of θ_1 using the delta method as shown in Wooldridge (1995a) and Wooldridge (2002, chapter 12).

For $y_{2it} = 1$, we define the general regressors for time period *t* as:

$$\widehat{w}_{it} = (W_{1it}, y_{3it}, 0, ..., 0, \hat{\lambda}_{it1}, 0, ..., 0, 0, ..., 0, \hat{\lambda}_{it2}, 0, ..., 0, 0, ..., 0, \hat{\lambda}_{it3}, 0, ..., 0, 0, ..., 0, \hat{\lambda}_{it4}, 0, ..., 0,)$$

and the parameter vector in the second stage is:

$$\theta_2 = (\gamma'_1, \alpha_1, \eta_{11}, \dots, \eta_{T1}, \eta_{12}, \dots, \eta_{T2}, \eta_{13}, \dots, \eta_{T3}, \eta_{14}, \dots, \eta_{T4})'$$

which is a $G \times 1$ vector where $G = (1 + K_1 + L + 4T)$.

We can write $E[\log(y_{1it})|w_{it}, y_{2it} = 1] = w_{it}\theta_2$, then we have: $\log(y_{1it}) = w_{it}\theta_2 + \varepsilon_{it}$ where $E[\varepsilon_{it}|w_{it}, y_{2it} = 1] = 0$ $(t = \overline{1, T})$.

On the selected sample, our POLS estimator is:

$$\widehat{\theta}_{2} = \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{2it}w_{it}'w_{it}\right)^{-1} \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{2it}w_{it}'\log(y_{1it})\right),$$
(F.20)

$$\hat{\theta}_2 = \theta_2 + \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T y_{2it} w'_{it} w_{it}\right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T y_{2it} w'_{it} \varepsilon_{it}\right),$$
(F.21)

and it can be shown that:

$$\sqrt{N}(\hat{\theta}_2 - \theta_2) \xrightarrow{d} Normal(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}),$$
 (F.22)

where

$$\mathbf{A} = E\left(\sum_{t=1}^{T} y_{2it} w'_{it} w_{it}\right), \tag{F.23}$$

$$\mathbf{B} = Var(\mathbf{h}_i) = E(\mathbf{h}_i \mathbf{h}'_i) \text{ and } \mathbf{h}_i = \mathbf{s}_i - \mathbf{F}\mathbf{r}_i,$$
(F.24)

in which

$$\mathbf{s}_i = \sum_{t=1}^T y_{2it} w'_{it} \varepsilon_{it}, \tag{F.25}$$

$$\mathbf{F} = E\left(\sum_{t=1}^{T} y_{2it} w_{it}' \theta_2' \nabla_{\theta_1} w_{it}'(\theta_1)\right),\tag{F.26}$$

in which $\nabla_{\theta_1} w'_{it}(\theta_1)$ is a $G \times Q$ gradient of $w'_{it}(\theta_1)$ evaluated at $\hat{\theta}_1$ and \mathbf{r}_i is defined in the previous part.

To estimate $Avar(\hat{\theta}_2) = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}/N$, we obtain:

$$\widehat{\mathbf{A}} \equiv N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{2it} \widehat{w}'_{it} \widehat{w}_{it}, \qquad (F.27)$$

$$\widehat{\mathbf{F}} \equiv N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{2it} \widehat{w}'_{it} \widehat{\theta}'_2 \nabla_{\theta_1} w'_{it} (\widehat{\theta}_1) \right),$$
(F.28)

and for each $i = \overline{1, N}$.

$$\widehat{\mathbf{s}}_{i} \equiv \sum_{t=1}^{T} y_{2it} \widehat{w}_{it}' \widehat{\varepsilon}_{it}, \qquad (F.29)$$

in which $\hat{\varepsilon}_{it} = \log(y_{1it}) - \widehat{w}_{it} \hat{\theta}_2$, and

$$\widehat{\mathbf{h}}_i = \widehat{\mathbf{s}}_i - \widehat{\mathbf{F}} \widehat{\mathbf{r}}_i. \tag{F.30}$$

A consistent estimator of **B** is:

$$\widehat{\mathbf{B}} \equiv N^{-1} \sum_{i=1}^{N} \widehat{\mathbf{h}}_i \widehat{\mathbf{h}}'_i.$$
(F.31)

The asymptotic variance of $\hat{\theta}_2$ is estimated as:

$$\widehat{Avar(\hat{\theta}_2)} = \widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}/N, \tag{F.32}$$

and the asymptotic standard errors are obtained as the square roots of the diagonal elements of this matrix.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Abadie, Alberto. 2000. Semiparametric estimation of instrumental variable models for causal effects. Working Paper 260. National Bureau of Economic Research.
- Amemiya, Takeshi. 1978. The estimation of a simultaneous equation generalized probit model. *Econometrica* 46(5). 1193–1205.
- Amemiya, Takeshi. 1979. The estimation of a simultaneous equation tobit model. *International Economic Review* 20(1). 169–81.
- Angrist, Joshua D. 2001. Estimation of limited-dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19(1). 2–16.
- Angrist, Joshua D. & William N. Evans. 1998. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review* 88(3). 450 77.
- Arellano, Manuel & Olympia Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68(1). 29–51.
- Arellano, Manuel & Bo Honore. 2001. Panel data models: Some recent developments. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of econometrics*, vol. 5 Handbook of Econometrics, chap. 53, 3229–3296. Elsevier.
- Baltagi, Badi H. & Qi Li. 1991. A transformation that will circumvent the problem of autocorrelation in an error-component model. *Journal of Econometrics* 48(3). 385–393.
- Baltagi, Badi H. & Ping X. Wu. 1999. Unequally spaced panel data regressions with ar1 disturbances. *Econometric Theory* 15(06). 814–823.
- Becker, Gary S. & H. Gregg Lewis. 1973. On the interaction between the quantity and quality of children. *Journal of Political Economy* 81(2). S279–88.
- Ben-Porath, Yoram & Finis Welch. 1976. Do sex preferences really matter? *The Quarterly Journal* of *Economics* 90(2). 285 307.
- Bhargava, Alok & J. D. Sargan. 1983. Estimating dynamic random effects models from panel data covering short time periods. *Econometrica* 51(6). 1635–59.
- Bloom, David, David Canning, Günther Fink & Jocelyn Finlay. 2009. Fertility, female labor force participation, and the demographic dividend. *Journal of Economic Growth* 14(2). 79–101.
- Blundell, Richard W. & James L. Powell. 2004. Endogeneity in semiparametric binary response models. *Review of Economic Studies* 71. 655–679.
- Blundell, Richard W. & Richard J. Smith. 1989. Estimation in a class of simultaneous equation limited dependent variable models. *Review of Economic Studies* 56(1). 37–57.

- Bronars, Stephen G. & Jeff Grogger. 2001. The effect of welfare payments on the marriage and fertility behavior of unwed mothers: Results from a twins experiment. *Journal of Political Economy* 109(3). 529–545.
- Browning, Martin. 1992. Children and household economic behavior. *Journal of Economic Literature* 30(3). 1434–75.
- Cain, Glen G. & Martin D. Dooley. 1976. Estimation of a model of labor supply, fertility, and wages of married women. *Journal of Political Economy* 84(4). S179–99.
- Cameron, Colin A. & Pravin K. Trivedi. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1(1). 29–53.
- Card, David & Daniel G. Sullivan. 1988. Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica* 56(3). 497–530.
- Carrasco, Raquel. 2001. Binary choice with binary endogenous regressors in panel data: Estimating the effect of fertility on female labor participation. *Journal of Business and Economic Statistics* 19(4). 385–394.
- Chamberlain, Gary. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* 47(1). 225–38.
- Chamberlain, Gary. 1992. Sequential moment restrictions in panel data: Comment. *Journal of Business and Economic Statistics* 10(1). 20–26.
- Chay, Kenneth Y. & Dean Hyslop. 1998. Identification and estimation of dynamic binary response panel data models: Empirical evidence using alternative approaches.
- Das, Mitali. 2002. Estimators and inference in a censored regression model with endogenous covariates. Discussion papers. Columbia University.
- Das, Mitali. 2005. Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics* 124(2). 335 – 361.
- Eckstein, Zvi & Kenneth I. Wolpin. 1990. Estimating a market equilibrium search model from panel data on individuals. *Econometrica* 58(4). 783–808.
- Even, William E. 1987. Career interruptions following childbirth. *Journal of Labor Economics* 5(2). 255–77.
- Fleisher, Belton M. & Jr. Rhodes, George F. 1979. Fertility, women's wage rates, and labor supply. *American Economic Review* 69(1). 14–24.
- Giles, J. & I. Murtazashvili. 2010. A control function approach to estimating dynamic probit models with endogenous regressors, with an application to the study of poverty persistence in china.
- Greene, William H. 1997. Econometric analysis. New York: Macmillan 3rd edn.

- Gronau, Reuben. 1973. The intrafamily allocation of time: The value of the housewives' time. *American Economic Review* 63(4). 634–51.
- Hausman, Jerry A. 1978. Specification tests in econometrics. Econometrica 46(6). 1251-71.
- Heckman, James J. 1974. Effects of child-care programs on women's work effort. *Journal of Political Economy* 82(2). S136–S163.
- Heckman, James J. 1978a. Dummy endogenous variables in a simultaneous equation system. *Econometrica* 46(4). 931–59.
- Heckman, James J. 1978b. Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. In Manski C.E. & Daniel L. McFadden (eds.), *The econometrics of panel data* 30/31, 227–269. University of Chicago.
- Heckman, James J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1). 153–61.
- Heckman, James J. 1981a. Heterogeneity and state dependence. In *Studies in labor markets* NBER Chapters, 91–140. National Bureau of Economic Research, Inc.
- Heckman, James J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In Manski C.E. & Daniel L. McFadden (eds.), *Structural analysis of discrete panel data with econometric applications*, MIT press.
- Heckman, James J., Robert J. Lalonde & Jeffrey A. Smith. 1999. The economics and econometrics of active labor market programs. In O. Ashenfelter & D. Card (eds.), *Handbook of labor economics*, vol. 3, chap. 31, 1865–2097. Elsevier.
- Heckman, James J. & Thomas E. Macurdy. 1980. A life cycle model of female labour supply. *Review of Economic Studies* 47(1). 47–74.
- Heckman, James J. & Robert J. Willis. 1974. Estimation of a stochastic model of reproduction: An econometric approach. NBER Working Papers 0034 National Bureau of Economic Research, Inc.
- Heckman, James J. & Robert J. Willis. 1977. A beta-logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy* 85(1). 27–58.
- Honore, Bo E. 1993. Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics* 59(1-2). 35–61.
- Honore, Bo E. & Luojia Hu. 2001. Estimation of censored regression models with endogeneity.
- Honore, Bo E. & Luojia Hu. 2004. Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics* 122(2). 293–316.

- Honore, Bo E. & Ekaterini Kyriazidou. 2000. Panel data discrete choice models with lagged dependent variables. *Econometrica* 68(4). 839–874.
- Hsiao, Cheng. 1986. Analysis of panel data. Cambridge, MA: Cambridge University Press.
- Hyslop, Dean R. 1999. State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 67(6). 1255–1294.
- Jacobsen, Joyce P., James Wishart Pearce III & Joshua L. Rosenbloom. 1999. The effects of childbearing on married women's labor supply and earnings: Using twin births as a natural experiment. *Journal of Human Resources* 34(3). 449–474.
- Kim, Jungho & Arnstein Aassve. 2006. Fertility and its consequence on family labour supply. IZA Discussion Papers 2162 Institute for the Study of Labor (IZA).
- Kim, Kyoo II. 2006. Sample selection models with a common dummy endogenous regressor in simultaneous equations: A simple two-step estimation. *Economics Letters* 91(2). 280–286.
- Kyriazidou, Ekaterini. 2001. Estimation of dynamic panel data sample selection models. *Review* of *Economic Studies* 68(3). 543–72.
- Labeaga, Jose M. 1999. A double-hurdle rational addiction model with heterogeneity: Estimating the demand for tobacco. *Journal of Econometrics* 93(1). 49–72.
- Lee, Lung-fei. 1999. Estimation of dynamic and arch tobit models. *Journal of Econometrics* 92(2). 355–390.
- Lee, Myoung-jee. 1996. Methods of moments and semiparametric econometrics for limited dependent variable models. Springer.
- Lehrer, Evelyn L. 1992. The impact of children on married women's labor supply: Black-white differentials revisited. *Journal of Human Resources* 27(3). 422–444.
- Mullahy, J. 1997. Instrumental-variable estimation of count data models: Applications to models of cigarette smoking behavior. *Review of Economics and Statistics* 79. 586–93.
- Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica* 46(1). 69–85.
- Nakamura, Alice & Masao Nakamura. 1992. The econometrics of female labor supply and children. *Econometric Reviews* 11(1). 1–71.
- Nelson, Forrest & Lawrence Olson. 1978. Specification and estimation of a simultaneous-equation model with limited dependent variables. *International Economic Review* 19(3). 695–709.
- Newey, Whitney K. 1985. Semiparametric estimation of limited dependent variable models with endogenous explanatory variables. *Annales de l'inséé* 59/60.
- Newey, Whitney K. 1986. Linear instrumental variable estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 32(1). 127–141.

- Newey, Whitney K. 1987. Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics* 36(3). 231–250.
- Newey, Whitney K. & Daniel L. McFadden. 1994. Large sample estimation and hypothesis testing. In Robert F. Engle & Daniel L. McFadden (eds.), *Handbook of econometrics*, vol. 4, chap. 36, 2111 – 2245. Elsevier.
- Nguyen, Hoa B. 2010. Estimating a fractional response model with a count endogenous regressor and an application to female labor supply. In William H. Greene & R. Carter Hill (eds.), *Advances in econometrics*, vol. 26, 253–298. Emerald Group Publishing Limited.
- Papke, Leslie E. & Jeffrey M. Wooldridge. 1996. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11(6). 619 – 32.
- Papke, Leslie E. & Jeffrey M. Wooldridge. 2008. Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics* 145(1-2). 121 133.
- Rivers, Douglas & Quang H. Vuong. 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39(3). 347–366.
- Rosenzweig, Mark R. & Kenneth I. Wolpin. 1980. Life-cycle labor supply and fertility: Causal inferences from household models. *Journal of Political Economy* 88(2). 328–48.
- Schultz, T. Paul. 1978. Fertility and child mortality over the life cycle: Aggregate and individual evidence. *American Economic Review* 68(2). 208–15.
- Semykina, Anastasia & Jeffrey M. Wooldridge. 2010. Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics* 157(2). 375–380.
- Shaw, Kathryn. 1994. The persistence of female labor supply: Empirical evidence and implications. *Journal of Human Resources* 29(2). 348–378.
- Skrondal, Anders & Rabe-Hesketh Sophia. 2004. *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models.* Boca Raton, FL: Chapman and Hall, CRC.
- Smith, Richard J. & Richard W. Blundell. 1986. An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica* 54(3). 679–85.
- Staiger, Douglas & James H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65(3). 557 – 586.
- Terza, Joseph V. 1998. Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics* 84(1). 129 154.
- Vella, Francis. 1993. A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review* 34(2). 441–57.
- Vella, Francis & Marno Verbeek. 1999. Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics* 90(2). 239–263.

- Vytlacil, Edward. 2002. Independence, monotonicity, and latent index models: An equivalence result. *Econometrica* 70(1). 331–341.
- Vytlacil, Edward & Nese Yildiz. 2007. Dummy endogenous variables in weakly separable models. *Econometrica* 75(3). 757–779.
- Weiss, Andrew A. 1999. A simultaneous binary choice/count model with an application to credit card approvals. In R. Engle & H. White (eds.), *Cointegration, causality, and forecasting: A Festschrift in honour of Clive W. J. Granger*, 429 – 461. Oxford and New York: Oxford University Press.
- Willis, Robert J. 1973. A new approach to the economic theory of fertility behavior. *Journal of Political Economy* 81(2). S14–64.
- Winkelmann, Rainer. 2000. Econometric analysis of count data. Berlin: Springer.
- Wooldridge, Jeffrey M. 1997. Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory* 13(5). 667–678.
- Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
- Wooldridge, Jeffrey M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20(1). 39–54.
- Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press 2nd edn.