This is to certify that the
dissertation entitled

THREE ESSAYS ON NONLINEAR MODELS FOR FRACTIONAL RESPONSE VARIABLES WITH TIME-VARYING INDIVIDUAL HETEROGENEITY

presented by

YOUNG GUI KIM

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Economics

Jeffrey M. Wooldridge

Major Professor's Signature

July 24, 2009

Date

THREE ESSAYS ON NONLINEAR MODELS FOR FRACTIONAL RESPONSE VARIABLES WITH TIME-VARYING INDIVIDUAL HETEROGENEITY

By

Young gui Kim

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Economics

2009

ABSTRACT

THREE ESSAYS ON NONLINEAR MODELS FOR FRACTIONAL RESPONSE VARIABLES WITH TIME-VARYING INDIVIDUAL HETEROGENEITY

By

Young gui Kim

This dissertation consists of three chapters on nonlinear panel models for fractional response variables. The first chapter extends the framework of Papke and Wooldridge (2008) by allowing unobserved individual heterogeneity to vary over time. An interaction term between a time effect and an individual effect of Kiefer (1980) and Lee (1991), known as an interactive effect, is added to the model. This allows each cross-sectional individual to respond differently to time effects common to all individuals. Based on Papke and Wooldridge (2008), I discuss the pooled Quasi maximum likelihood estimator (QMLE) and the multivariate weighted nonlinear least squares (MWNLS) estimator to estimate the model in strictly exogenous case. The two-step procedure of Rivers and Vuong (1988) and the single-step estimator of Wooldridge (2007) are discussed in the endogenous case. The methods are applied to analyze the effect of school spending on math test pass rates of 4th graders in Michigan. According to the Monte-Carlo simulations, the fractional time-varying model gives the least root mean squared errors among all three models in a reasonable range of correlation between the regressors and individual heterogeneity.

The second chapter continues studying fractional response models and considers binary endogenous explanatory variables (EEVs). Because EEVs are not continuous, I modify the bivariate probit model to derive a conditional mean function for a fractional response

variable. The pooled quasi-limited information maximum likelihood estimator (QLIMLE) is proposed based on this mean function. The conditional mean function for a fractional response variable and the reduced form for a binary endogenous variable are estimated together to identify all parameters of interest [Wooldridge (2007)]. Also the average treatment effects (ATEs) of a binary EEV are discussed. The estimation method is applied to study the effect of fertility on women's fractions of working hours. The simulations show the root mean squared errors of the ATEs of the fractional bivariate model are less than both the linear and the fractional probit model.

In the last chapter, I discuss the hurdle model for a fractional dependent variable. Binary endogenous explanatory variables and time-varying individual heterogeneity are considered as before. The hurdle model allows us to separate the determination of corner solutions (y=0 or y>0) from the decision of the amount of dependent variables conditional on y>0. This hurdle model allows us to use different sets of EEVs for the corner solution equation and the equation for the amount and to obtain the ATEs conditional on y>0. I study the fertility effect of the second chapter again by using the hurdle model.

ACKNOWLEDGMENTS

I cannot express enough thanks to my advisor, Professor Jeffrey M. Wooldridge for his enthusiastic advice and support. Without his advice and discussion, I could not have written this dissertation. He also provided me the data for the first chapter. I am very grateful to Professor Timothy Vogelsang. His generous support made me possible to successfully finish the doctoral program. I would like to thank my committee members, Professor Leslie E. Papke and Professor Yuehua Cui for their valuable comments and constructive discussions. I would like to express graditute to my family for their understanding and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1 FRACTIONAL RESPONSE MODELS WITH TIME-VARYING INDIVIDUAL HETEROGENEITY

1.1 Introduction

Some economic variables are measured as fractions, percentages, or proportions. Examples include market shares, 401(k) participation rates, test pass rates, expenditure shares, and cost shares. Many models and estimation methods have been proposed for fractional dependent variables. Fractional variables always takes values between 0 and 1 including two extreme values. A simple linear model is easy to be estimated and interpreted, but the fitted values could be greater than 1 or negative even though the dependent variable is defined in the unit interval. One might use the log-odd transformation to avoid this problem, but we may not recover the original dependent variable without further assumptions.

Papke and Wooldridge (1996, 2008) proposed a new approach to modeling fractional response variables based on a correctly specified conditional mean function. I focus on their two estimation methods. The first one is the pooled quasi-maximum likelihood estimator (QMLE) based on the Bernoulli distribution. With large enough samples, models can consistently estimated without distributional assumption by their estimation methods [For detailed discussion, see Papke and Wooldridge (1996), Gourieroux, Monfort, and Trognon (1984)]. The second method is the weighted non-linear least squares (WNLS) estimator for a non-linear mean function. A multivariate WNLS (MWNLS) estimator is proposed to address serial correlations in the cross sectional dimensions of their panel data.

In this chapter, I combine this framework with the time-varying individual effects of Kiefer (1980) and Lee (1991). In most panel analyses, unobserved individual heterogeneity has been assumed as a time-invariant factor. Kiefer (1980) and Lee (1991) introduced time-varying individual heterogeneity in a linear model by adding interaction terms between individual effects and time effects. I use this interaction term (sometimes called an interactive effect) to relax the constant individual heterogeneity assumption in our non-linear model[1]. This allows each individual to respond differently to the common time effects. One example where this can be useful is in a model of income where the productivity of an individual (unobserved ability) varies over the business cycle [Ahn, Lee and Schmidt (2007)]. In our application, the effect of school spendings on math test pass rates in Michigan is studied by using school district level data. The model with time-varying individual heterogeneity allows each school district to respond differently to the difficulty levels of the math tests.

The remainder of the chapter is organized as follows: In section 2, I introduce the basic model. Section 3 and section 4 contain estimation methods under the assumption of strictly exogenous explanatory variables and endogenous explanatory variables, respectively. I provide three test statistics for the null hypothesis of no time-varying individual effect in section 5. I apply the estimation methods to the test pass rates for fourth graders in Michigan and conduct Monte-Carlo simulations in sections 6 and section 7. The last section concludes.

1.2 Basic Model

Let us assume N randomly drawn cross-sectional units from the population and T

[1] While Lee (1991), and Ahn, Lee, Schmidt (2007) considered mulitple time-varying factors in a linear model, one time-varying factor in a nonlinear model is considered because of an identification problem.

observations for each cross-sectional observation. The number of cross-sectional units is assumed to be substantially larger than the number of time series observations (large N fixed T asymptotics).

Let's assume a correctly specified conditional mean function:

$$E\left(y_{it}|x_{it}, c_i\right) = \Phi\left(x_{it}\beta + \eta_t c_i\right), \tag{1.1}$$

where i indexes a cross-sectional unit and t indexes time ($i = 1, 2, ..., N,\ t = 1, 2, ..., T$). y_{it} denotes a fractional dependent variable for individual i and time t that is continuous between 0 and 1 (also y_{it} is allowed to take 0 or 1 with positive probability), and x_{it} is a $1 \times K$ vector of explanatory variables. c_i indicates unobserved individual heterogeneity (an individual effect) and η_t is its corresponding coefficient that is allowed to vary over time. η_t can be regarded as a time effect in that this coefficient is common to all cross-sectional units at time t. Therefore, $\eta_t c_i$ is sometimes called an interactive effect [Bai (2005)]. This term is assumed to be linear additive in the model. To ensure that the fitted values take values between 0 and 1, cumulative distribution functions (cdf) can be used for the mean function of y_{it}. In particular, the standard normal cdf, $\Phi(\cdot)$, is used in this chapter because this gives a computationally simple estimator when the model contains unobserved heterogeneity and endogenous regressors.

The main contribution of this paper is to allow the effect of unobserved individual heterogeneity to vary over time in the linear index. One might attempt to consider the most flexible individual effect which varies freely over time (c_{it}). However, there are two problems: First, it is impossible to distinguish a time-varying individual effect (c_{it}) from an idiosyncratic error (u_{it}) without further assumptions. Second, it seems unreasonable to assume that individual heterogeneity, such as innate ability, changes over time. An alternative is an interaction term between a time effect (η_t) and a constant individual effect (c_i). We can interpret this interactive effects in two ways: the effect of individual heterogeneity is varying over time or time dummies have individual specific intercepts. This interactive effect model includes the constant individual effect model as a special case and still retains a manageable structure. I set $\eta_1 = 1$ as a normalization.

Because this is a nonlinear model, we are interested in the partial effects not the coefficients. Dropping subscript i, we can obtain the effect of the kth regressor, $x_{t,k}$, on y_t from $\partial E(y_t|x_t, c)/\partial x_{t,k} = \beta_k \phi(x_t\beta + \eta_t c)$ if $x_{t,k}$ is continuous, or $\Phi(x_t^{(1)}\beta + \eta_t c) - \Phi(x_t^{(0)}\beta + \eta_t c)$ where $x_t^{(1)}$ and $x_t^{(0)}$ denote two different values of $x_{t,k}$ if $x_{t,k}$ is discrete. As we can see, the partial effect depends on x_t and c. It is a good idea to evaluate this effect at interesting values of x's or average the partial effects across individuals. However, we may not observe c. Therefore, a popular method is to calculate the average partial effects (APEs) by averaging the partial effects across the distribution of c. The APEs can be obtained from $E_c[\beta_k \phi(x_t\beta + \eta_t c)]$ or $E_c[\Phi(x_t^{(1)}\beta + \eta_t c) - \Phi(x_t^{(0)}\beta + \eta_t c)]$.

In order to identify both β and APEs, we use two assumptions:

$$E\left(y_{it}|x_i, c_i\right) = E\left(y_{it}|x_{it}, c_i\right), \tag{1.2}$$

where $x_i \equiv (x_{i1}, ..., x_{iT})$ is a set of explanatory variables in all time periods. The first assumption is that x_{it} conditional on c_i is strictly exogenous. This assumption rules out dynamic models and feedback models. Also this implies that we do not consider traditional simultaneous models and the omitted variable problem. If appropriate instrumental variables are available, this assumption can be relaxed so that explanatory variables are allowed to be endogenous.

The second assumption is

$$c_i = \psi + \bar{x}_i\xi + a_i, a_i|x_i \sim N\left(0, \sigma_a^2\right), \tag{1.3}$$

where $\bar{x}_i \left(\equiv T^{-1}\sum_{t=1}^{T} x_{it}\right)$ is a $1 \times K$ vector of time-averaged explanatory variables and a_i denotes the part of unobserved individual heterogeneity which is not correlated with x_i. σ_a^2 is the variance of c_i given x_i. This Mundlak (1978)-Chamberlain (1980) device is adopted to restrict the distribution of c_i conditional on x_i, which enables us to get the APEs by using the law of iterated expectations (LIE). This implies that constant individual heterogeneity is correlated with time-invariant parts of strictly exogenous explanatory variables. For the purpose of flexibility, one can consider more general functional forms for $D(c_i|x_i)$ by adding squared and/or interaction terms. Also parametric modelling for $Var(a_i|x_i)$ or relaxing the normality assumption of $a_i|x_i$ can be considered, but I focus on (1.3) in this chapter. [For detailed discussion about more general specifications for D($c_i|x_i$), see Papke and Wooldridge (2008).]

Under assumptions (1.2) and (1.3),

$$\begin{aligned} E\left(y_{it}|x_i, a_i\right) &= \Phi\left(\eta_t\psi + x_{it}\beta + \eta_t\bar{x}_i\xi + \eta_t a_i\right) \qquad (1.4) \\ \eta_t a_i|x_i &\sim N\left(0, \eta_t^2\sigma_a^2\right). \end{aligned}$$

x_{it} must contain only time-varying elements so that there is no perfect multicollinearity between x_{it} and $\bar{x}_i$. Adding time dummies to the model is desirable in order to allow different intercept for each time. Even though $\eta_t\psi$ is absorbed into the coefficients of time dummies, I do not consider time dummies explicitly for notational simplicity.

By the properties of the normal distribution, we can rewrite the equation (1.4) as

$$\begin{aligned} E(y_{it}|x_i) &= \Phi\left(\frac{\eta_t\psi + x_{it}\beta + \eta_t\bar{x}_i\xi}{\sqrt{1+\eta_t^2\sigma_a^2}}\right) \\ &\equiv \Phi\left(\psi_{at} + x_{it}\beta_{at} + \bar{x}_i\xi_{at}\right), \end{aligned} \tag{1.5}$$

where $\psi_{at}\left(\equiv \frac{\eta_t\psi}{\sqrt{1+\eta_t^2\sigma_a^2}}\right)$, $\beta_{at}\left(\equiv \frac{\beta}{\sqrt{1+\eta_t^2\sigma_a^2}}\right)$, and $\xi_{at}\left(\equiv \frac{\eta_t\xi}{\sqrt{1+\eta_t^2\sigma_a^2}}\right)$ denote scaled parameters. The subscript a means that all parameters are scaled by σ_a^2 and the subscript t means those parameters depend on time due to η_t. Because we can observe $(1, x_{it}, \overline{x}_i)$, the scaled parameters $(\psi_{at}, \beta_{at}, \xi_{at})$ are identified. Papke and Wooldridge (2008) proposed methods to estimate parameters and APEs consistently without distributional assumptions about $D(y_{it}|x_i, c_i)$.

By the LIE, the average structural function is obtained as follows.

$$E_{\bar{x}_i}\left[\Phi(\psi_{at} + x_t\beta_{at} + \bar{x}_i\xi_{at})\right]. \tag{1.6}$$

The APEs for each t can be obtained by differentiating equation (1.6) with respect to $x_{t,k}$. $\bar{x}_i$ is redundent in the conditional mean function and c_i and x_{it} are independent conditional on $\bar{x}_i$. With these two assumptions, we can use the arguments in Wooldridge (2002, sectional 2.2.5). Therefore, the APEs of x_k at time t can be consistently estimated by $N^{-1}\sum_{i=1}^{N}\hat{\beta}_{at,k}\phi\left(\hat{\psi}_{at} + x_{it}\hat{\beta}_{at} + \bar{x}_i\hat{\xi}_{at}\right)$ for each t. A scale factor also can be obtained from $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi\left(\hat{\psi}_{at} + x_{it}\hat{\beta}_{at} + \bar{x}_i\hat{\xi}_{at}\right)$. We can compare coefficients of the nonlinear models with those of linear models after multiplying the coefficients of the nonlinear model by this scale factor.

1.3 Estimation methods with strictly exogenous variables

Let us define $w_{it} \equiv (1, x_{it}, \overline{x}_i)$ and $\pi_t \equiv (\psi_{at}^{'}, \beta_{at}^{'}, \xi_{at}^{'})'$ for notational simplicity. Then, the conditional mean function is $\Phi(w_{it}\pi_t)$. Before proceeding, I need to discuss two practical issues: normalization of σ_a^2 and recovering parameters of interest. For identification, σ_a is set to 1 as normalization, but this does not affect the results because we have a nonlinear model. From now on, I will drop the subscript a. Our interest lies in $\theta \equiv (\psi', \eta', \beta', \xi')'$ not $\pi \equiv \left(\pi_1^{'}, \pi_2^{'}, ..., \pi_T^{'}\right)^{'}$, and there are specific relationships between π and θ (For example, $\psi_t = \psi / \sqrt{1 + \eta_t^2}$). Therefore, two approaches can be used: the classical minimum distance (CMD) estimator and estimation with constraints. First, the CMD estimator recovers θ from π by minimizing the weighted differences in the relationships between π and θ. This CMD estimator uses the variance-covariance matrix for $\hat{\pi}$ as weight. The detailed explanation for the CMD estimator will be provided later. Second, one can estimate the model after imposing the relationships directly on the objective functions. Estimation methods with constraints are straightforward.

Now focus on the ways to estimate π consistently. Given the conditional mean function, there are many possible estimators. One estimator is the pooled nonlinear least squares (PNLS) estimator. This estimator is consistent and $\sqrt{N}$-asymptotically normal (T is fixed). However, the pooled NLS estimator is inefficient because this assumes homoskedasticity, but this assumption does not hold. A reasonable model for the variance, $V(y_{it}|x_i)$, has the following form:

$$V\left(y_{it} \mid x_i\right) = \tau\Phi\left(w_{it}\pi_t\right)\left[1 - \Phi\left(w_{it}\pi_t\right)\right], \tag{1.7}$$

where τ is a constant. [Readers are referred to Papke and Wooldridge (1996) for detailed discussion.] Based on (1.7), the weighted nonlinear least squares (WNLS) estimator can be an alternative. In this chapter, the pooled quasi-maximum likelihood estimator (QMLE) based on Bernoulli distribution is used instead of the WNLS estimator. The pooled QMLE is equivalent to the WNLS when the inverse of (1.7) is used as a weighting matrix , but the pooled QMLE is simpler because the NLS estimator needs preliminary results to obtain the weighting matrix. Also this estimator is strongly consistent even if the true distribution of y is not Bernoulli once the first moment is assumed to be correctly specified because the Bernoulli distribution is belong to the linear exponential family (LEF) [Papke and Wooldridge (1996), Gourieroux, Monfort and Trognon (1984)]. The pooled QMLE is obtained by maximizing the sum of the following log-likelihood function. I call this estimator the pooled fractional probit estimator (FPE) as Papke and Wooldridge (2008) did.

$$\begin{aligned} \ell_{it}\left(w_{it};\pi\right) &= y_{it}\log\left[\Phi(\psi_t + x_{it}\beta_t + \bar{x}_i\xi_t)\right] \\ &\quad +(1-y_{it})\log\left[1-\Phi(\psi_t + x_{it}\beta_t + \bar{x}_i\xi_t)\right] \\ &\equiv y_{it}\log\left[\Phi(w_{it}\pi_t)\right] + (1-y_{it})\log\left[1-\Phi(w_{it}\pi_t)\right]. \end{aligned} \tag{1.8}$$

Another possible estimator to enhance the efficiency is the multivariate weighted nonlinear least squares (MWNLS) estimator which considers serial correlations among errors within a cross sectional unit i. Let define $w_i \equiv (D, Dx_{it}, D\overline{x}_i)$ which is a $T \times (T + 2TK)$ matrix of explanatory variables. $D \equiv I_{T\times T}$ denotes a set of all time dummies. Dx_{it} denotes $T \times TK$ matrix which has x_{it} for its diagonal elements and 0 for all off-diagonal elements. In the same way, $D\overline{x}_i$ is defined as $T \times TK$ matrix with its diagonal elements of $\overline{x}_i$. This matrix contains all interaction terms between $\overline{x}_i$ and D. Suppose the conditional mean function for $y_i \equiv (y_{i1}, y_{i2}, ..., y_{iT})'$ be $m_i(w_i, \pi) \equiv \Phi(w_i\pi)$. Applying MWNLS estimator requires the variance-covariance matrix of y_i, $Var(y_i|x_i)$. Along with (1.7), we need to form $Cov(y_{it}, y_{is}|x_t), t \neq s$. Instead of using a parametric model for $Var(y_i|x_i)$, its working version proposed in the generalized estimating equations (GEE) literature is used. Also the standardized errors are assumed to have the same correlation (exchangeable correlation assumption). The common correlation is calculated by $\tilde{\rho} = [NT(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{t \neq s} \tilde{e}_{it}\tilde{e}_{is}$, where $\tilde{e}_{it} = \tilde{u}_{it}/\sqrt{\Phi(w_{it}\tilde{\pi}_t)\left[1-\Phi(w_{it}\tilde{\pi}_t)\right]}$ and $\tilde{u}_{it} \equiv y_{it} - \Phi(w_{it}\tilde{\pi}_t)$. '$\sim$' means that all are evaluated at the preliminary consistent estimates from the pooled FPE. The MWNLS estimator solves

$$\min_{\theta} \sum_{i=1}^{N} [y_i - m_i(w_i, \pi)]' \tilde{V}_i^{-1} [y_i - m_i(w_i, \pi)] \tag{1.9}$$

$$\tilde{V}_i = Diag\left[\tilde{\Phi}_i(1-\tilde{\Phi}_i)\right]^{1/2} C(\tilde{\rho}) Diag\left[\tilde{\Phi}_i(1-\tilde{\Phi}_i)\right]^{1/2},$$

where $\tilde{V}_i$ is the $T \times T$ matrix where its diagonal elements are from (1.7) and its off-diagonal elements from the 'working correlation matrix'. $C(\tilde{\rho})$ is the $T \times T$ matrix where its diagonal elements are 1 and its off-diagonal elements are $\tilde{\rho}$ [Papke and Wooldridge (2008), Liang and Zeger (1986)]. This MWNLS estimator is equivalent to the GEE when the same weighting matrix is used. Note that we cannot consider the serial correlations when we recover the parameters by the CMD estimator because pure cross-section data is available for each t. Therefore only constrained MWNLS (or constrained GEE) can be used to consider possible serial correlations.

For testing hypotheses and using the CMD estimator, we need a consistent estimator of asymptotic variance matrix. Hypotheses include independence between the unobserved heterogeneity and regressors ($H_0 : \xi = 0$) and time-invariant individual heterogeneity ($H_0 : \eta_t = 1$). Because we do not assume that the conditional variance of y is correctly specified, a consistent estimator of asymptotic variance robust to heteroskedasticity and serial correlation has so-called sandwich form [Arellano (1987), Wooldridge (2002)].

With regularity conditions, the pooled FPE is $\sqrt{N}$-asymptotically normal with the variance of

$$\widehat{Avar}\left(\hat{\pi}_{QMLE}\right) = \hat{A}^{-1}\hat{B}\hat{A}^{-1}/N, \tag{1.10}$$

where $\hat{A} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(-\nabla_{\pi}^{2}\ell_{it}(\hat{\pi})\right)$, $\hat{B} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t\neq s}\widehat{s}_{is}\widehat{s}_{it}'$, and $\widehat{s}_{it} = \nabla_{\pi}\ell_{it}(\widehat{\pi})$ [Wooldridge (2002, Theorem 12.3)].

In the case of the MWNLS estimator has the following asymptotic variance.

$$\begin{aligned}\widehat{Avar}(\hat{\pi}_{MWNLS}) &= \left(\sum_{i=1}^{N}\nabla_{\pi}\hat{m}_i'\hat{V}_i^{-1}\nabla_{\pi}\hat{m}_i\right)^{-1} \\ &\quad\times\left(\sum_{i=1}^{N}\nabla_{\pi}\hat{m}_i'\hat{V}_i^{-1}\hat{u}_i\hat{u}_i'\hat{V}_i^{-1}\nabla_{\pi}\hat{m}_i\right) \\ &\quad\times\left(\sum_{i=1}^{N}\nabla_{\pi}\hat{m}_i'\hat{V}_i^{-1}\nabla_{\pi}\hat{m}_i\right)^{-1}/N,\end{aligned} \tag{1.11}$$

where $\nabla_{\pi}\hat{m}_i$ is the Jacobian of the mean function, $\hat{m}_i$ is $m_i(w_i, \pi)$ evaluated at $\hat{\pi}$, and $\hat{u}_i(\equiv y_i - m_i(w_i, \hat{\pi}))$ is a residual vector.

With scaled estimates ($\hat{\pi}$) and their asymptotic variances ($\hat{\Xi}$), the CMD estimator is used to estimate θ from $\hat{\pi}$. We have $(T + 2K)$ parameters of interest (θ), but $\hat{\pi}$ is a

$(T + 2TK) \times 1$ vector. Let's suppose $h(\cdot)$ be a continuously differentiable function to represent relationships between θ and π ($\pi = h(\theta)$). The CMD estimator minimizes the distance between $\hat{\pi}$ and $h(\hat{\theta})$.

$$\min_{\theta} \left[\hat{\pi} - h(\theta)\right]' \widehat{\Xi}^{-1} \left[\hat{\pi} - h(\theta)\right]. \tag{1.12}$$

From the first order condition for $\hat{\theta}$ and a standard mean value expansion around θ, the following asymptotic distribution of $\hat{\theta}$ is obtained.

$$\sqrt{N}\left(\hat{\theta} - \theta\right) \overset{a}{\sim} N\left(0, H(\theta)'\Xi^{-1}H(\theta)\right), \tag{1.13}$$

where $H(\theta)$ ($\equiv \nabla_{\theta} h(\theta)$) is the $(T + 2TK) \times (T + 2K)$ Jacobian of $h(\theta)$. The appropriate estimator of asymptotic variance of $\hat{\theta}$ is

$$\widehat{Avar(\hat{\theta})} = (\hat{H}'\hat{\Xi}^{-1}\hat{H})^{-1}/N, \tag{1.14}$$

where $\hat{H}$ is H evaluated at $\hat{\theta}$. When we use the estimators with constraints to get $\hat{\theta}$ directly, the robust asymptotic variance of $\widehat{\theta}$ has the same form with (1.14).

1.4 Estimation methods with endogenous explanatory variables

Now let's relax the strict exogeneity assumption given appropriate instrumental variables. Time-varying individual heterogeneity is considered as above. One endogenous variable case is considered in this chapter, but considering multiple endogenous variables case is straightforward. The basic model with an endogenous explanatory variable is as follows:

$$E\left(y_{it1}|y_{it2}, z_{it}, c_{i1}, v_{it1}\right) = E\left(y_{it1}|y_{it2}, z_{it1}, c_{i1}, v_{it1}\right) \tag{1.15}$$
$$= \Phi\left(\alpha_1 y_{it2} + z_{it1}\delta_1 + \eta_t c_{i1} + v_{it1}\right)$$

where y_{it2} denotes an endogenous explanatory variable which is assumed to be correlated not only with individual heterogeneity (c_{i1}) but also with unobserved omitted variable (v_{it1}). z_{it} ($\equiv (z_{it1}, z_{it2})$) denotes a set of exogenous explanatory variables and instrumental variables. Instrumental variables (z_{it2}) should not have significant effects on y_{it1} and must be at least partly correlated with y_{it2}.

The Mundlak(1978) -Chamberlain(1980) device is applied again ($c_{i1}|z_i \sim N\left(\psi_1 + \bar{z}_i\xi_1 + a_{i1}, \sigma_{a1}^2\right)$).

$$E(y_{it1}|y_{it2}, z_i, r_{it1}) = \Phi(\eta_t\psi_1 + \alpha_1 y_{it2} + z_{it1}\delta_1 + \eta_t\bar{z}_i\xi_1 + r_{it1}) \tag{1.16}$$

where r_{it1} denotes a composite error of $\eta_t a_{i1}$ and v_{it1}. In this structural equation, y_{it2} is assumed to be correlated with r_{it1}.

The reduced form for y_{it2} is assumed as

$$y_{it2} = \psi_{2t} + z_{it}\delta_2 + \bar{z}_i\xi_2 + r_{it2} \equiv w_{it2}\gamma + r_{it2}. \tag{1.17}$$

This equation can be consistently estimated by the pooled OLS estimator which gives the same estimates with the fixed effects estimator because the averages of explanatory variables are added. Endogeneity of y_{it2} implies that r_{it2} is correlated with both a_{i1} and v_{it1}. I assume that r_{it1} given r_{it2} is conditionally normal and this leads $r_{it1} = \rho_t r_{it2} + e_{it}, e_{it}|z_i \sim N(0,\ \tau_t^2)$.

Substituting r_{it1} with $\rho_t r_{it2} + e_{it}$,

$$
\begin{aligned}
E(y_{it1}|y_{it2}, z_i, r_{it2}) &= \Phi(\eta_t\psi_1 + \alpha_1 y_{it2} + z_{it1}\delta_1 + \eta_t\bar{z}_i\xi_1 \qquad (1.18)\\
&\quad +\rho_t r_{it2} + e_{it})\\
&= \Phi\left(\frac{\eta_t\psi_1 + \alpha_1 y_{it2} + z_{it1}\delta_1 + \eta_t\bar{z}_i\xi_1 + \rho_t r_{it2}}{\sqrt{1+\tau_t^2}}\right).
\end{aligned}
$$

Because joint normality of (r_{it1}, r_{it2}) is assumed, the final model (1.18) comes from properties of normal distribution. τ_1 is set to 1 as normalization.

With a continuous endogenous variable, two possible approaches can be used to estimate the equation (1.18) consistently: the two-step estimation procedure and the limited information maximum likelihood estimator(LIMLE). The LIMLE is called the single step estimator compared to the two-step estimation procedure. The two-step procedure proposed by Rivers and Vuong (1988) is as follows.

1. Estimate the reduced form for y_2 (1.17) and get the residuals, $\widehat{r}_{it2}$.

2. Replace r_{it2} with $\widehat{r}_{it2}$ as an additional regressor and estimate the structural model (1.18).

Now all explanatory variables are observable and the omitted variable problem can be avoided by adding $\widehat{r}_{it2}$. Even though $\widehat{r}_{it2}$ is added to the model, y_{it2} is still allowed not to be strictly exogenous. Therefore, I use the FPE instead of the MWNLS estimator. The MWNLS estimator requires the strict exogeneity as the generalized least squares (GLS) estimator in linear models does. The endogeneity ($H_0 : \varphi_t = 0$) can be tested by using the Wald statistics. If the null hypothesis is rejected, we should adjust the asymptotic variance for inference because the final model (1.18) contains a generated regressor ($\widehat{r}_{it2}$) from the first step estimation. In other words, the asymptotic variance of $\sqrt{N}\left(\widehat{\theta}-\theta\right)$ depends on the asymptotic variance of $\sqrt{N}\left(\widehat{\gamma}-\gamma\right)$ [For detailed discussion, Wooldridge (2002) ch. 12.4 and Papke and Wooldridge (2008) Appendix 1].

This two-step procedure has some advantages. Along with computationally simplicity, the endogeneity can be easily tested because we do not have to adjust asymptotic variances under the null hypothesis. However, the asymptotic variances should consider generated regressors if the exogeneity is rejected, and this approach cannot be used if endogenous explanatory variables are not continuous.

Wooldridge (2007) proposed the quasi limited information MLE (QLIMLE) of nonlinear models without additional assumptions of the joint distribution of y_1 and y_2. Substituting r_{it2} with $(y_{it2}-w_{it2}\gamma)$, we can rewrite the equation (1.18) as

$$\begin{aligned} E(y_{it1}|y_{it2},z_i,w_{it2}) &= \Phi[\frac{\eta_t\psi_1+\alpha_1 y_{it2}+z_{it1}\delta_1+\eta_t\bar{z}_i\xi_1}{\sqrt{1+\tau_t^2}} && (1.19) \\ &\quad +\frac{\rho_t\left(y_{it2}-w_{it2}\gamma\right)}{\sqrt{1+\tau_t^2}}] && (1.20) \\ &\equiv \Phi(w_{it1}\pi). \end{aligned}$$

Because I assume a correctly specified mean function and the Bernoulli is the member of the linear exponential family, the Bernoulli QMLE estimates the equation (1.19) consistently. To identify all parameters, equation (1.19) is estimated along with the equation (1.17). The the quasi-log-likelihood function is as follows:

$$\begin{aligned} l_{it}(\theta) &= y_{it}\log\left[\Phi(w_{it1}\pi_t)\right]+(1-y_{it})\log\left[1-\Phi(w_{it1}\pi_t)\right] && (1.21) \\ &\quad -(1/2)\log(\sigma_{r2}^2)-(1/2)\left[(y_{it2}-w_{it2}\gamma)^2/\sigma_{r2}^2\right], \end{aligned}$$

where $\sigma_{r2}^2\equiv Var(r_{it2}|w_{it2})$. Compared to the two-step procedure, the QLIMLE might be more efficient because both the structural equation and the reduced form for the endogenous variable are estimated together, and estimating valid standard errors is straightforward. Also this approach can be used when the models contain discrete or limited endogenous variables, while the two-stage estimator is invalid [Wooldridge (2007)].

1.5 Test statistics for time-varying individual effects

In this section, I discuss test statistics for testing constant individual heterogeneity ($H_0 : \eta_t = 1$). Let us start from the basic estimating equation of Papke and Wooldridge (2008), $E(y_{it}|x_i) = \Phi(\psi + x_{it}\beta + \overline{x}_i\xi)$. To consider time-varying individual heterogeneity, consider a more general estimating equation, $E(y_{it}|x_i) = \Phi(\psi + x_{it}\beta + \overline{x}_i\xi + \delta_t(\overline{x}_i\xi \times T_t)) \equiv \Phi(w_{it}\pi + \delta_t(\overline{x}_i\xi \times T_t)) \equiv m_{it}$, where $\overline{x}_i\xi \times T_t$ indicates interaction terms between $\overline{x}_i\xi$ and all time dummies. The degree of freedom does not depend on the number of regressors in this case.

The first statistic is a regression-based test which is also known as a variable addition test. This is a special case of Ramsey's RESET in that fitted values are used as additional explanatory variables. The procedure is as follows:

1. Estimate the null model and get the coefficients of $\overline{x}_i$ $\left(\widetilde{\xi}\right)$.
2. Add interaction terms $\left(\overline{x}_i\widetilde{\xi} \times T_t\right)$ between $\overline{x}_i\widetilde{\xi}$ and time dummies.
3. Estimate the model of y_{it} on 1, a set of all time dummies, $x_{it}, \overline{x}_i, \left(\overline{x}_i\widetilde{\xi} \times T_t\right)$.
4. Test whether all coefficients of the interaction terms are 0 ($H_0 : \delta_t = 0$).

This test can be implemented by using usual statistical packages such as E-views or STATA. Also it is easy to make the test statistics robust to possible misspecified second moments.

The score or Lagrangian multiplier (LM) statistic is ideally suited for specification testing. Only restricted estimates are needed to implement the score test. In most cases, restricted models are quite easier to estimate. Let the restriction be $c(\theta) = 0$, $(\theta \equiv (\pi', \delta_t')')$ and its Jacobian matrix be C. With consistent estimates, each element of the statistic is evaluated at the restricted parameters, $\widetilde{\theta} = (\tilde{\pi}', 0')'$. Because I do not assume correctly specified variance, the general form of the statistics should be used.

$$LM = \frac{1}{N}\left(\sum_{i=1}^{N} S_i(\widetilde{\theta})\right)' \widetilde{A}^{-1}\widetilde{C}'\left[\widetilde{C}\widetilde{A}^{-1}\widetilde{B}\widetilde{A}^{-1}\widetilde{C}'\right]^{-1}\left(\sum_{i=1}^{N} S_i(\widetilde{\theta})\right) \overset{a}{\sim} \chi^2_{(T-1)}, \tag{1.22}$$

where $S_i(\widetilde{\theta})$ is the score function evaluated at the restricted estimates. A and B are defined in (1.10), respectively. the LM statistic should be 0 evaluated at the unrestricted estimates $(\widehat{\theta})$. If the restriction is valid, the LM statistic evaluated at the restricted estimates $(\widetilde{\theta})$ also should be close to 0. The following procedure can used to produce the same test statistic with (1.22). Divide θ into the $(K - T + 1)\times 1$ vector of π and the $(T - 1)\times 1$ vector of δ_t. Let $\nabla_\pi m_{it}$ and $\nabla_{\delta_t} m_{it}$ denote the gradients with respect to π and δ_t.

1. Run a multivariate regression $\nabla_{\delta_t} m_{it}$ on $\nabla_\pi m_{it}$ and obtain the $1 \times (T - 1)$ residuals, $\tilde{r}_{it}$. Then make $\tilde{u}_{it}\tilde{r}_{it}$ by multiplying $\tilde{u}_{it}$ by each elements of $\tilde{r}_{it}$, where $\tilde{u}_{it}$ is the residuals from the restricted model.

2. Run the regression 1 on $\tilde{u}_{it}\tilde{r}_{it}$ and construct $LM = NT - SSR$, where SSR is the usual sum of squared residuals. This statistic has a limiting χ^2_{T-1} distribution.

The last one is the minimum distance statistic which comes from the CMD estimator. Going back to the time-varying model, let's suppose a vector of the restricted parameters be $\alpha \equiv (\psi', 1, \beta', \xi')'$, where $\eta_t = 1$ is imposed and $g(\cdot)$ shows the relationship between π

and α $[\pi = g(\alpha)]$. The minimum distance statistic is

$$N\left[(\widehat{\pi} - g(\widehat{\alpha}))\,\widehat{\Xi}^{-1}\,(\widehat{\pi} - g(\widehat{\alpha}))\right] - N\left[\left(\widehat{\pi} - h(\widehat{\theta})\right)'\widehat{\Xi}^{-1}\left(\widehat{\pi} - h(\widehat{\theta})\right)\right] \overset{a}{\sim} \chi^2_{T-1}. \quad (1.23)$$

This statistic is based on the distance between two minimized objective functions. The intuition of the statistics is the same with that of likelihood ratio (LR) test. If the restriction is true, the distance must be close to 0. However, we need to get $\widehat{\theta}$ and $\widehat{\alpha}$ to obtain this statistic like the LR test.

1.6 Empirical application: math test pass rates

In 1994, Michigan reformed K-12 school funding system in order to equalize educational opportunities in terms of school spending. From a policy prospective, it is important to estimate the effect of school spending on student performance consistently. Papke (2005, 2008) found that increase in per-student spending has significantly positive effect on student performance measured as pass rates on fourth grade math test (the Michigan Educational Assessment Program, MEAP, test). In the paper, she allowed the endogeneity of spending and used foundation grant as an instrument variable for school spending. While Papke (2008) adopted the fixed effects estimator to control school heterogeneity in the linear model, Papke and Wooldridge (2008) provided a new approach to consider a nonlinear conditional mean function with unobserved heterogeneity. They also found a significantly positive effect of school spending on math test pass rates. I use the same data with Papke and Wooldridge (2008). Table (A.1) provides simple descriptive statistics for key variables over the years of 1995 through 2001 used in this chapter. Papke and Wooldridge (2008) used 501 school districts, while 503 districts are used here.

Equation (1.24) shows the mean function for math test pass rates with time-varying unobserved school district heterogeneity.

$$
\begin{aligned}
E(math4_{it}|\cdot) \;=\; & \Phi[T_t\delta_t + \beta_1 \log(avgrexpp)_{it} + \beta_2 lunch_{it} \\
& +\beta_3 \log(enroll)_{it} + \xi_1 \overline{\log(avgrexpp)}_i + \xi_2 \overline{lunch}_i \\
& +\xi_3 \overline{\log(enroll)}_i + \eta_t c_i],
\end{aligned} \tag{1.24}
$$

where $math4_{it}$ denotes the fraction of 4th graders who pass the MEAP fourth grade math test in district i for year t. The sample includes 503 school districts and 7 years. T_t is a set of time dummies to allow a different intercept for each year. Papke and Wooldridge (2008) used three explanatory variables: the 4-year averaged value of spending per pupil in real dollar ($\log(avgrexpp)$), the fraction of students who are eligible for free or reduced lunch program ($lunch$), and the number of enrollment ($\log(enroll)$). Because Papke (2005) found that not only the current spending but lagged spendings have significant effects on the performance, averaged real spending per pupil in first, second, third, and fourth grade ($avgrexpp$) is considered. Logarithm transformation implies a diminishing effect of spending and enrollment on the test pass rates. The purpose of this model is to find the effect of spending on the test pass rates, so all variables which might be correlated with spending and have significant effect on the test pass rates should be controlled. Two control variables are added into the model; $lunch$ is used as a proxy for the poverty rate or economic well-being, and $\log(enroll)$ is for controlling school size.

Our estimation procedure is as follows: (1) estimate equation (1.24) for each year by using the pooled FPE and get total 49 parameters ($\widehat{\pi}$). There are 7 parameters and each school district has 7 years of observations. (2) Recover the parameters of interest

$\theta = (\delta', \eta', \beta', \xi')$ by the CMD estimator from $\widehat{\pi}$. Also the model with the restriction of $\pi = h(\theta)$ is estimated by both the pooled constrained FPE and the constrained MWNLS estimator.

Table (A.2) and (A.3) contain the estimates of Papke and Wooldridge (2008)'s and a linear model. Because all models are nonlinear, the APEs of all three explanatory variables are provided to compare these results with those of the linear model. The first column of the table (A.2) contains the results of the fixed effects estimator of the linear model, and the next columns are the replicated results of Papke and Wooldridge (2008). The table (A.3) shows the estimation results of the time-varying model (the model with time-varying individual heterogeneity). Three estimation methods are used: the CMD estimator after the pooled FPE for each time, the constrained FPE (the constrained fractional probit estimator), and the constrained MWNLS estimator. The constrained estimators indicate the estimators after imposing the relationships between π and θ on the conditional mean function.

The results show two things. First, the effect of school spending on the performance is still significant and positive after considering time-varying individual heterogeneity. The coefficients of $\log(avgrexpp)$ of the time-varying model (0.949 ~1.103) are slightly bigger than those of Papke and Wooldridge (2008) (0.883 ~0.886), but it is meaningless to compare the coefficients directly. Comparing the APEs, I find the effect of the time-varying model (0.214 ~0.228) is smaller than those of Papke and Wooldridge (2008) (0.298). Under the time-varying heterogeneity assumption, increase in spending per student by 10 percentage points leads increase in test pass rates by 2.1 to 2.2 percentage points. These APEs of the nonlinear models are smaller than the coefficient of $\log(avgrexpp)$ in the linear model.

Second, $\hat{\eta}_t$ varies from 0.84 to 1.24. Table (A.2) contains the results of the LM test and the variable addition test. The constant heterogeneity hypothesis is reject at 5% significant level. Also the minimum distance statistic in the table (A.3) is 35 and its p-value is zero to four decimal places. Because the LM test and the variable addition test are to test a necessary condition for the time-varying heterogeneity, minimum distance statistics is greater. These three test statistics show that the assumption of constant individual heterogeneity is too restrictive.

Under proposal A, each school district was given the foundation grant based on its spending level per student in 1994. As a result, all districts received per-student spending at least as much as a basic grant. For example, in 1995, school districts spent under $4200 in 1994 were given $4200 per student or an additional $250 per student. High revenue districts were held harmless in that they could receive at least as much amount as before. As discussed in Papke (2005) and Papke and Wooldridge (2008), the foundation grant variable is a non-smooth function of spending in 1994. After controlling for spending in 1994, it seems reasonable to assume that the foundation grant is not correlated with unobserved shocks that affect the test pass rate. On the other hand, it is easily show that the foundation allowance is partially correlated with the average expenditure per student because a district's revenue is largely determined by the foundation amount.

As discussed above, the estimating equation for math test pass rates has the contrl function form,

$$
\begin{aligned}
E(math4_{it}|\cdot) &= \Phi[T_t\delta_{t1} + \alpha\log(avgrexpp)_{it} + \delta_{11}lunch_{it} \qquad (1.25)\\
&\quad +\delta_{12}\log(enroll)_{it} + \xi_{11}\overline{lunch}_i + \delta_{13t}\log(rexpp_{94})_i \\
&\quad +\xi_{12}\overline{\log(enroll)}_i + \rho_t r_{it2} + e_{it}],
\end{aligned}
$$

where $\log(rexpp_{94})$ denotes logarithm of real expenditure per pupil in 1994. Adding this variable allows us to assume tha the foundation allowance is unrelated to the error, e_{it}. The variable r_{it2} is the reduced form error from the equation for $\log(avgrexpp)$ (1.26).

The reduced form equation is

$$
\begin{aligned}
\log(avgrexpp)_{it} &= T_t\delta_{t2} + \delta_{12}lunch_{it} + \delta_{21}\log(enroll)_{it} \qquad (1.26)\\
&\quad +\delta_{22t}\log(rexpp_{94})_i + \delta_{23t}\log(grant)_{it} + \xi_{21}\overline{lunch}_i \\
&\quad +\xi_{22}\overline{\log(enroll)}_i + r_{it2},
\end{aligned}
$$

where $\log(grant)_{it}$ denotes logarithm of real foundation grant. I have two estimation strategies. First, estimate equation (1.26) and get residuals ($\widehat{r}_{it2}$). Substitute r_{it2} in (1.25) with the residuals and estimate the model by the pooled FPE (and the CMD estimator) or the constrained FPE. Second, equation (1.26) and equation (1.25) can be estimated at the same time by the pooled Quasi-LIMLE (the single step estimator).

Table (A.4) contains three results: the first column shows the result of the fixed effects instrumental variable estimator, and the rest columns are the replicated results of Papke and Wooldridge (2008) using the two-step FPE and the quasi-LIMLE, respectively. School spending still has a significantly positive effect on the test pass rates, and the effects are much greater than those in the exogenous case. This implies that the effect of spending has downward bias under the exogeneity assumption. The QLIMLE has smaller standard errors and bigger APEs than the two-step procedure. The coefficient of spending in the linear model is between them. In both methods, t-statistics reject the null hypothesis of the exogeneity of spending. The hypothesis of constant individual heterogeneity is rejected at 5% significant level.

Table (A.5) and table (A.6) show the results of the fractional model with time-varying individual heterogeneity. Table (A.5) contains the results of the two-step procedure, while table (A.6) reports the results of the single-step estimator (QLIME). In both tables, the spending has a significantly positive effect on the math test pass rates. In the two-step procedure, both the CMD estimator and the constrained FPE report slightly different coefficients of the spending and the constrained FPE seems to be less efficient than the CMD estimator. In the single step estimator, both estimation methods give quite similar results and the constrained methods (CQLIMLE) looks more efficient. The APEs of the fractional model with time-varying individual heterogeneity are from 0.507 to 0.592, and are similar with the coefficient of the linear model. Under endogenous spending assumption, increase in school spending per student by 10 percent points results in increase in the math test pass rates by 5.6 percentage points. The exogeneity of spending is rejected at 5% level, and the constant individual heterogeneity is also rejected.

Table (A.7) shows the scale factors at different spending levels. Using linear models, Papke (2005) found different effects of spending at different spending levels and different initial performance levels, by estimating separate models for various subgroups. Papke and Wooldridge (2008) used a more parsimonious approach to allowing nonconstant partial effects by applying the fractional probit model, but with constant coefficients on the heterogeneity. The table contains the scale factors for nine cases: strictly exogenous and endogenous spending and the two-step procedure and the single-step estimator for endogenous spending. For each case, three results are reported: the fractional probit model of Papke and Wooldridge (2008), the CMD estimator and the FPE (or the CMWNLS estimator in exogenous case) of the time-varying model, for the earliest and latest years in the data, 1995 and 2001. From the results, we can see the effect of school spending is greater at school districts with lower spendings. Let us focus on the results of the QLIMLE for the time-varying model with endogenous spending because we already reject the time constant heterogeneity and exogeneity of spending. The APEs are 0.668 in 1995 and 0.526 in 2001 at 5th the percentile of spending, while the APEs are 0.456 and 0.327 in 1995 and 2001 at 95th percentile. We can find diminishing APEs of school spending as spending levels go up. The difference in the APEs in 1995 and 2001 of Papke and Wooldridge (2008) is 0.03 (0.827-0.797). When we allow time-varying heterogeneity, the difference becomes 0.142 (0.668-0.526). Generally, there are nontrivial differences in the APEs across different parts of the spending distribution when time-varying factors are allowed on the unobserved heterogeneity.

1.7 Monte-Carlo simulations

We can study the small-sample properties of the various estimation methods, when the models are and are not correctly specified, via a simulation study. Because some of the models are time consuming to estimate, I use 500 Monte Carlo replications. Nine different sample size combinations are used: the number of cross-sectional units are 100, 300, and 500 and the number of time periods are 5, 7, and 10.

Equation (1.27) shows the data generating process of an explanatory variable.

$$
\begin{aligned}
x_{it} &= \gamma_{xc} c_i + \sqrt{1-\gamma_{xc}^2} e_{it} \\
c_i, e_{it} &\sim N(0,1),
\end{aligned} \tag{1.27}
$$

where x_{it} denotes an explanatory variable and is allowed to be correlated with unobserved heterogeneity (c_i). The variance of x_{it} is set to 1.

The data generating process of a fractional response variable is as follows. H Beroulli outcomes, w_{ith} are generated with probabilities that depend on the covariate, the unknown heterogeneity, and unobservables u_{it}. The observed fractional response is the fraction of successes.

$$
\begin{aligned}
p_{it} &= p(w_{ith}=1) = \Phi(\gamma_{yt} T_t + \beta x_{it} + \eta_t c_i + u_{it}), u_{it} \sim N(0,1) \\
y_{it} &= \frac{1}{H} \sum_{h=1}^{H} w_{ith},\ w_{ith} \sim Bernoulli(p_{it}),
\end{aligned} \tag{1.28}
$$

where T_t is time dummies. The data generating mechanism is the same as creating the fractional variable, y_{it}, from the Binomial distribution in which the probability of a success is defined as equation (1.28) and there are H trials. I use $H = 1000$.

The normalization I use on the time-varying coefficients is $\eta_1 = 1$. I set up two different populations: η_t is increasing by 0.1 and by -0.1 over time. γ_{yt} is $T \times 1$ vector of known numbers and I set $\beta = 1$. Let the correlation between x and c be ρ_{xc}. Because variances of x and c are set to 1, the correlation of x and c is the same with the coefficient, γ_{xc}. Nine different correlations between x and c from 0.1 to 0.9 are considered. The larger ρ_{xc} is, the better the APEs of the fractional model with time-varying heterogeneity fit intuitively.

Only exogenous case is considered because both the endogeneity effect and time-varying individual heterogeneity might be mixed up so that we have difficulty in interpreting results. Six estimators are compared. The first one is the fixed effects estimator for the purpose of comparison. The next two are the pooled FPE and the MWNLS estimator proposed by Papke and Wooldridge (2008). The last three are the CMD estimator based on estimates of the pooled FPE for each time, the constrained FPE (CFPE), and the constrained MWNLS (CMWNLS) estimator.

Because the model in the population is nonlinear, a closed form for the average paratial effects is difficult to obtain. In stead, the APEs in the population are approximated as

$$\begin{aligned} APE_{t,pop} &= E_{(c,u)}\left[\beta\phi\left(\gamma_{yt}T_t + \beta x_t + \eta_t c + u_t\right)\right] \\ &\approx \beta\frac{1}{N}\sum_{i=1}^{N}\phi\left(\frac{\gamma_{yt}T_t + \beta x_{it}}{\sqrt{1+\eta_t^2}}\right), \end{aligned} \tag{1.29}$$

that is, I use a sample average across the distribution of x_{it} but with the true values of the parameters inserted.

Equation (1.30) shows the conditional mean function and the estimated APE:

$$\begin{aligned} E(y_{it}|x_i) &= \Phi(r_t T_t + \beta_t x_{it} + \delta_t \bar{x}_i) \\ APE_t &= \hat{\beta}_t\frac{1}{N}\sum_{i=1}^{N}\phi(\hat{r}_t T_t + \hat{\beta}_t x_{it} + \hat{\delta}_t \bar{x}_i), \end{aligned} \tag{1.30}$$

where, of course, these will differ by estimation method. To indicate different estimation methods, I write $APE_{t,l}$.

By comparing the root mean squared errors (RMSE) and biases between the APEs of the population and those of all estimators, I attempt to find the effect of considering time-varying individual heterogeneity on the APEs.

$$RMSE_l = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(APE_{t,l} - APE_{t,pop})^2} \tag{1.31}$$

Table (A.8) through table (A.14) contain the simulation results. Two cases are considered as explained above: an increasing individual effect and a decreasing individual effect. Table (A.8) through (A.11) show the RMSEs and biases from ρ_{xc} =0.1 to 0.9 when N is 500. The results shows the RMSEs become bigger as ρ_{xc} increases. When ρ_{xc} is large ($\rho_{xc} > 0.5$), all estimators have quite similar RMSEs. In large ρ_{xc}, the time-varying model gives slightly larger RMSEs than even a linear model. Note that high correlations between x and c mean there is not enough variations in x, which results in imprecise estimates and large RMSEs. So I restrict ρ_{xc} to reasonable ranges (from 0.2 to 0.4). Table (A.12) through (A.14) report the results of different N and different T. In most cases, the time-varying model gives better results than not only the linear model but the model of Papke and Wooldridge (2008). The constrained MWNLS estimator, which accounts, albeit in a misspecified way, for serial correlation, generally shows the best results. For examples, when $\rho_{xc} = 0.4$ and η_t is increasing, the RMSEs of the time-varying model are 0.0208~0.0203 with $N = 500$ and $T = 10$. The RMSE of the linear model is 0.0270, and those of Papke and Wooldridge (2008) are 0.0270~0.0271. If we calculate the APEs at different percentiles, Papke and Wooldridge (2008) model must show lower RMSEs than the linear model.

1.8 Conclusion

I allow unobserved individual heterogeneity to vary over time in a fractional response model. Based on the framework of Papke and Wooldridge (2008), the model is extended by adding an interaction term (an interactive effect) between a time effect and an individual effect proposed by Kiefer (1980) and Lee (1991). This interaction term can be interpreted as a time-varying coefficient of individual heterogeneity or an individual specific intercept at each time. Therefore, each individual is allowed to respond differently to the common effects (the time effect). This setting contains the constant individual effect model as a special case and we still have a manageable form. The Mundlak-Chamberlain device is used to restrict of the distribution of individual effect conditional on explanatory variables. Based on a conditional mean function for fractional dependent variables, the pooled fractional probit estimator and the multivariate non linear least squares estimator are discussed as in Papke and Wooldridge (2008). Two cases are considered: strictly exogenous explanatory variables and continuous endogenous variables with appropriate instrumental variables. Especially, I adopt the single-step estimator of Wooldridge (2007) along with the traditional two-step estimator when the model contains endogenous explanatory variables. Also I mention three test statistics for time-varying individual heterogeneity. In particular, the variable addition test can be conducted by using usual statistical software programs such as the STATA even though the time-varying model cannot be estimated by those kinds of software.

I apply the estimation methods to analyze the effects of school spending per student on students' performance in terms of the math test pass rates in school district level. The time-varying individual heterogeneity model allows each school district to respond differently to common time effects such as the difficulty level of the test. I find that the effect of school spending on test pass rates is still statistically significant after considering time varying individual heterogeneity and the APEs are quite similar in both the time-varying model and the time-constant heterogeneity model. However, when the

APEs are evaluated at different percentiles of spending and in different years, the APEs of the time-varying model are smaller that those of the time-constant model. The hypothesis of a constant individual effect is rejected by all three test statistics.

The Monte-Carlo simulations are also conducted to find finite sample properties of estimators. Two cases are considered: an increasing and a decreasing individual effect using nine different samples with N=100, 300, 500 and T=5, 7, 10. Also different correlations between the regressor and individual heterogeneity from 0.1 to 0.9 are considered. When our interest is restricted to the reasonable ranges of the correlations between an explanatory variable and individual heterogeneity (0.2~0.4), the time varying model gives less RMSEs than both the linear model and the model with constant individual effects. Especially, constrained MWNLS estimator which considers a possible serial correlation shows the best results among all estimator.

Chapter 2 THE FRACTIONAL MODELS WITH BINARY ENDOGENOUS EXPLANATORY VARIABLES AND TIME-VARYING INDIVIDUAL HETEROGENEITY

2.1 Introduction

In this chapter, I continue to study the fractional response model, but now I allow for the presence of a binary endogenous explanatory variable. I combine the problem of an endogenous binary explanatory variable with unobserved heterogeneity. As in the first chaper, I allow time-varying coefficients on the individual heterogeneity.

In the first chapter, continuous EEVs are considered and an extension of the two-step control function approach proposed by Rivers and Voung (1998) is used. However, it is well known that there is no consistent two-step estimator when EEVs are not continuous in nonlinear models. Instead, the bivariate probit model is modified to derive a conditional mean function and the single step estimator. Recently, Wooldridge (2007) has shown that in some cases – fractional responses being a leading one – quas-limited information maximum likelihood can be used. Therefore, the main contribution of this chapter is to derive the conditional mean function for fractional dependent variables with binary EEVs and to extend Wooldridge's approach to the panel data case.

The model here has many applications. For example, y_{it1} could be the proportion of questions correctly answered on a test, and y_{it2} an intervention, such as getting individual tutoring help. Or, y_{it1} can be the fraction of pension assets invested in the stock market and y_{it2} and indicator for taking a "financial awareness" class. The empirical example I use here is to analyze the effect of fertility on female labor supply. Women are still primarily responsible for providing child care and doing housework. Jacobsen (1994) pointed out that women spend four times as much time for child care as men. Therefore, it is not surprising that many studies has found a negative relationship between fertility and women's labor supply. However, to interpret this causal relationship clearly, we have to consider the endogeneity of fertility in female labor supply equation. Using my model and estimation method, I can address two kinds of endogeneity: fertility and female labor supply are likely to be jointly determined and unobserved individual heterogeneity might affect both variables.

To consider this endogeneity, simultaneous equation systems may be used to consider joint determination of fertility and labor supply. However, if there is a misspecification in any equation, this misspecification transmits its inconsistency to all system. In other words, even if our interest lie in the labor supply equation and this model is correctly specified, inconsistent estimates might be obtained when the other model has a misspecification problem. An alternative is to use an instrumental variable estimation method with appropriate instruments for fertility in reduced form for labor supply. Rosenzweig and Wolpin (1980) used twins in the first birth as an instrument. Because twining occurs randomly, this would produce consistent estimates, but we have a sample size problem. There are only 87 twin mothers in Rosenzweig and Wolpin (1980). Angrist and Evans (1998) proposed a new instrumental variable, an indicator for whether the sexes of the first two children are the same, for fertility based on the parental preference on the mixed sex sibling in developed country[2]. Angrist and Evans (1998) compared two instrumental

[2] Park and Cho (1995), and Arnold and Zhaoxiang (1981) use parental preference for boys in developing countries.

variable results: twins at the first birth and the same sex. This allows them to find the effect of children of different ages. Carrasco(2001) extended Angrist and Evans (1998) to panel data, but he focuses on the same sex. Some studies use lagged dependent variables as proxies to control unobserved heterogeneity, while panel data allow us to control individual heterogeneity explicitly. Instead adding lagged dependent variables, I use Mundlak(1978) - Chamberlain (1980) device to restrict the distribution of unobserved heterogeneity.

Another issue is definitions of labor supply and children variables. There are two major measures of labor supply variables: job participation (extensive margin) and working hours (intensive margin). Angrist and Evans (1998) used participation, weeks worked, and hours per a week, while Carrasco's (2001) dependent variable is labor force participation. The effect of fertility on labor supply is not clear a prior because an income effect and a substitution effect coexist. However, in most cases mothers are likely to withdraw entirely from labor market or at least reduce their working hours to take care of their new born babies. The binary dependent models might underestimate the effect of the fertility in that the binary indicator does not change as long as mothers take part in the labor market even if they reduce working hours. In this paper, fraction of working hours to total available hours is used as a dependent variable. Many researchers use linear models for working hours, but working hours have two limits: maximum time limit and 0. Therefore, linear models cannot ensure the fitted values exist between these limits. Also constant effects of independent variables might be too restrictive. In this paper, I focus on the effect of new born babies because most of the effect of fertility depends on that more than, for example, on the number of children [Carrasco (2001)].

2.2 The Model with binary endogenous explanatory variables

As in the first chapter, I assume N randomly drawn cross-sectional units from the population and T observations for each cross-sectional observation. Again, large N and fixed T asymptotics is used.

Let us assume a correctly specified conditional mean function as

$$E(y_{it1}|y_{it2}, z_{it1}, c_{i1}, v_{it1}) = \Phi(\alpha y_{it2} + z_{it1}\beta_1 + \eta_t c_{i1} + v_{it1}), \qquad (2.1)$$

where i indexes a cross-sectional unit and t indexes time ($i = 1, 2, ..., N, t = 1, 2, ..., T$). y_{it1} denotes a fractional dependent variable which can take on any value between zero and one, including zero and one. The variable y_{it2} indicates a binary endogenous explanatory variable (EEV). In this chapter, I focus on one EEV case; in principle, it one can extend the approach to more than one EEV. z_{it1} is a set of strictly exogenous regressors. The interactive effect, $\eta_t c_i$, is added to consider time-varying individual heterogeneity again. v_{it1} is an unobserved omitted variable and is thought to be correlated with y_{it2}. Also the standard normal cumulative distribution function (cdf) is used for the mean function of y_{it1} to make the expected value between zero and one. This particular choice also provides some computational advantages.

Our main interest lies in the partial effect of a binary EEV. We can obtain this from $\Phi(\alpha + z_{t1}\beta_1 + \eta_t c_1 + v_{t1}) - \Phi(z_{t1}\beta_1 + \eta_t c_1 + v_{t1})$ dropping the subscript i. The partial effect depends on z_{t1}, c_1, and v_{t1}[3]. We can evaluate the effect at interesting values of z's or average the effects across individuals, but we may not observe c_1 and v_{t1}. Therefore, a popular method is to calculate the average treatment effects (ATEs) by averaging the

[3] This partial effect is called treatment effect when the variable of interst is binary or discrete.

treatment effect across the distribution of c_1 and v_1. Therefore,

$$ATE_t = E_{(c_1,v_1)}[\Phi(\alpha + z_{t1}\beta_1 + \eta_t c_1 + v_{t1}) - \Phi(z_{t1}\beta_1 + \eta_t c_1 + v_{t1})]. \tag{2.2}$$

To identify the parameters and the ATEs, the Mundlak(1978)-Chamberlain(1980) device is adopted to restrict the distribution of individual heterogeneity (c_1).

$$c_{i1}|z_i \backsim N(\psi_1 + \bar{z}_i\zeta_1 + a_{i1}, \sigma_{a1}^2), \tag{2.3}$$

where $z_i \equiv (z_{i1}, ..., z_{iT})$, $\bar{z}_i \equiv 1/T\sum_{t=1}^{T} z_{it}$, and $\sigma_{a1}^2 \equiv Var(a_{i1}|z_i)$. $z_{it} \equiv (z_{it1}, z_{it2})$ is a set of all exogenous variables and z_{it2} indicates a set of instrumental variables.

Rewrite the mean function with the device as follows.

$$E(y_{it1}|y_{it2}, z_i, r_{it1}) = \Phi(\eta_t\psi_1 + \alpha y_{it2} + z_{it1}\beta_1 + \eta_t\bar{z}_i\zeta_1 + r_{it1}), \tag{2.4}$$

where $r_{it1} \equiv (\eta_t a_{i1} + v_{it1})$ denotes a composite error. The endogeneity of y_{it2} implies that r_{it1} is correlated with y_{it2}.

To consider the nonlinearity of y_2, I use the probit model. There is no theoretical difference between the probit and logit model. However, because plugging probit fitted values for y_{it2} into the structural model is not valid, we exploit the properties of the normal distribution. (I refer to Wooldridge (2002) in the cross section case for probit.)

The equation (2.5) shows the reduced form model for y_{it2}:

$$\begin{aligned} E(y_{it2}|z_{it}, c_{i2}) &= \Phi(z_{it}\beta_2 + c_{i2} + v_{it2}) \\ c_{i2}|z_i &\backsim N(\psi_2 + \bar{z}_i\zeta_1 + a_{i2}, \sigma_{a2}^2), \end{aligned} \tag{2.5}$$

where c_{i2} indicates individual heterogeneity but is assumed invariant over time for simplicity. v_{it2} is unobservable but assumed independent of z_{it} and c_{i2} for identification. σ_{a1}^2 is the conditional variance of a_{i2}. With the Mundlak (1978) - Chamberlain (1980) device,

$$\begin{aligned} E(y_{it2}|z_i, r_{it2}) &= \Phi(\psi_2 + z_{it}\beta_2 + \bar{z}_i\zeta_2 + r_{it2}) \\ &= \Phi(w_{it2}\pi_2 + r_{it2}),\ w_{it2} \equiv (1, z_{it}, \bar{z}_i), \end{aligned} \tag{2.6}$$

where $r_{it2} \equiv (a_{i2} + v_{it2})$ is defined to represent a composite error and is independent of z_{it}. w_{it2} and π_2 are defined for notational simplicity. Because the endogeneity of y_{it2} implies that (r_{it1}, r_{it2}) are correlated, a linear projection of r_{it1} on r_{it2} and joint normality are assumed as: $r_{it1} = \rho_t r_{it2} + \varepsilon_{it}$, $\varepsilon_{it}|\,(z_i, r_{it2}) \sim N(0, \tau_t^2.)$. ρ_t is allowed to vary over time because r_{it1} contains a time-varying factor, the coefficient. From joint normality of (r_{it1}, r_{it2}) and independency of ε_{it} on y_{it2} and z_i,

$$\begin{aligned} E(y_{it1}|y_{it2}, z_{it}, r_{it2}) &= \Phi\left(\frac{\eta_t\psi_1 + \alpha y_{it2} + z_{it1}\beta_1 + \eta_t\bar{z}_i\zeta_1 + \rho_t r_{it2}}{\sqrt{1+\tau_t^2}}\right) \\ &\equiv \Phi(\eta_t\psi_{1\tau} + \alpha_\tau y_{it2} + z_{it1}\beta_{1\tau} + \eta_t\bar{z}_i\zeta_{1\tau} + \rho_{t\tau} r_{it2}) \\ &\equiv \Phi(w_{it1}\pi_{1t\tau} + \rho_{t\tau} r_{it2}), \end{aligned} \tag{2.7}$$

where subscript τ means all parameters are scaled by τ. For notational simplicity, $w_{it1} \equiv (1, y_{it2}, z_{it1}, \bar{z}_i)$ and its corresponding coefficient vector, $\pi_{1t\tau}$ are defined and I drop subscript τ from now on.

z_i and r_2 conditional on c_1 and v_1 are redundant in equation (2.1) and c_1, r_1, and y_2 are independent conditional on (z_i, r_2). By the law of iterated expectations,

$$
\begin{aligned}
ATE_t &= E_{(\bar{z}_i, r_2)}[\Phi\left(\eta_t\psi_1 + \alpha + z_{t1}\beta_1 + \eta_t\bar{z}_i\zeta_1 + \rho_t r_{it2}\right) \\
&\quad -\Phi\left(\eta_t\psi_1 + z_{t1}\beta_1 + \eta_t\bar{z}_i\zeta_1 + \rho_t r_{it2}\right)].
\end{aligned} \tag{2.8}
$$

From the properties of the normal distribution, we can rewrite (2.8) as

$$
\begin{aligned}
ATE_t &= E_{\bar{z}_i}[\Phi\left(\frac{\eta_t\psi_1 + \alpha + z_{t1}\beta_1 + \eta_t\bar{z}_i\zeta_1}{\sqrt{1+\rho_t^2}}\right) \\
&\quad -\Phi\left(\frac{\eta_t\psi_1 + z_{t1}\beta_1 + \eta_t\bar{z}_i\zeta_1}{\sqrt{1+\rho_t^2}}\right)].
\end{aligned} \tag{2.9}
$$

2.3 Estimation methods

The models with binary EEVs may not be estimated by the traditional two-step estimation, where generated regressors in the first step estimation are added as additional explanatory variables or used proxy variables for EEVs. In the nonlinear settings, this two-step estimator produces inconsistent estimates because we cannot pass the expectation through nonlinear functions. An alternative is to use the full information maximum likelihood estimator (FIMLE) after assuming the joint distribution of (y_1, y_2) given z. However, the FIMLE produces inconsistent estimates if the assumed joint distribution is not true in the population or there is misspecification in any equation.

In this chapter, the bivariate probit model is modified to obtain a conditional mean function for y_1 and the Bernoulli quasi maximum likelihood estimator (QMLE) is used based on the conditional mean function. So, I call this the fractional bivariate probit estimator (FBPE). To get the conditional mean function, take expectation with respect to r_2.

$$E(y_{it1}|y_{it2}, z_i) = \tag{2.10}$$

$$y_{it2}E[\Phi(w_{it1}\pi_{1t} + \rho_t r_{it2})|r_{it2} > -w_{it2}\pi_2]$$

$$+(1 - y_{it2})E[\Phi(w_{it1}\pi_{1t} + \rho_t r_{it2})|r_{it2} < -w_{it2}\pi_2]$$

Because r_2 is assumed normally distributed,

$$\begin{aligned} E(y_{it1}|y_{it2}, z_i) &= \frac{y_{it2}}{\Phi(w_{it2}\pi_2)} \int_{-\infty}^{w_{it2}\pi_2} \Phi(w_{it1}\pi_{1t} + \rho_t r_2)\phi(r_2)dr \\ &+ \frac{1 - y_{it2}}{\Phi(-w_{it2}\pi_2)} \int_{-\infty}^{-w_{it2}\pi_2} \Phi(w_{it1}\pi_{1t} + \rho_t r_2)\phi(r_2)dr \\ &= \frac{\Phi_2[w_{it1}\pi_{1t}, (2y_{it2} - 1)w_{it2}\pi_{2t}, (2y_{it2} - 1)\rho_t]}{\Phi[(2y_{it2} - 1)w_{it2}\pi_{2t}]} \equiv m_{it}. \end{aligned} \tag{2.11}$$

To identify all parameters, the structural equation (2.11) and the reduced form (2.6) can be estimated together. Wooldridge (2007) argues that, because the Bernoulli log likelihood is in the linear exponential family, only the mean $E(y_{it1}|y_{it2}, z_i)$ needs to be correctly specified along with the probit model for y_{it2} conditional on z_i. Here I am just applying a pooled method, and so the argument is essentially identical. The quasi-log-likelihood function is

$$\begin{aligned} \ln \ell_{it} &= y_{it1} \ln(m_{it}) + (1 - y_{it1}) \ln(1 - m_{it}) \\ &+ y_{it2} \ln[\Phi(w_{it2}\pi_2)] + (1 - y_{it2}) \ln[1 - \Phi(w_{it2}\pi_2)] \end{aligned} \tag{2.12}$$

where $m_{it} = E(y_{it1}|y_{it2}, z_i)$.

With consistent estimates, the ATEs of y_{it2} can be obtained from equation (2.13).

$$\widehat{ATE}_t = \frac{1}{N}\sum_{i=1}^{N}[\Phi\left(\frac{\hat{\eta}_t\hat{\psi}_1 + \hat{\alpha} + z_{t1}\hat{\beta}_1 + \hat{\eta}_t\bar{z}_i\hat{\zeta}_1}{\sqrt{1+\hat{\rho}_t^2}}\right) \tag{2.13}$$
$$-\Phi\left(\frac{\hat{\eta}_t\hat{\psi}_1 + z_{t1}\hat{\beta}_1 + \hat{\eta}_t\bar{z}_i\hat{\zeta}_1}{\sqrt{1+\hat{\rho}_t^2}}\right)].$$

The endogeneity of y_{it2} can be easily tested by the two step estimator of the Rivers-Vuoung (1988). If some of the ρ_t are different from zero, we have to use the FBP estimator to account for endogeneity of y_{it2}. If ρ_t is zero, the conditional mean function can be reduced to the simple mean function without endogeneity, that is

$$\frac{\Phi_2[w_{it1}\pi_{1t}, (2y_{it2}-1)w_{it2}\pi_{2t}, 0]}{\Phi[(2y_{it2}-1)w_{it2}\pi_{2t}]} \equiv \frac{\Phi[(w_{it1}\pi_{1t}]\Phi[(2y_{it2}-1)w_{it2}\pi_{2t}]}{\Phi[(2y_{it2}-1)w_{it2}\pi_{2t}]} \equiv \Phi[(w_{it1}\pi_{1t}].$$

For inference, we need a consistent estimator for the asymptotic variance.

$$Avar(\pi) = A^{-1}BA^{-1}/N, \tag{2.14}$$

where $A = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} -E[\nabla_\pi^2 \ln \ell_{it}(\pi)]$, $B = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} s_{is}(\pi)s_{it}(\pi)'$, and $s_{it}(\pi) \equiv \nabla_\pi \ln \ell_{it}(\pi)$ [Wooldridge (2002) Theorem 12.3]. The gradient of the quasi-log-likelihood function for each (i,t), $\nabla_\pi \ln \ell_{it}(\pi)$, can be written as

$$\begin{aligned}
\frac{\partial \ln \ell_{it}}{\partial \pi_{1t}} &= \frac{y_{it1}-m_{it}}{m_{it}(1-m_{it})}\frac{\phi(w_{it1}\phi\pi_{1t})\phi(v_{it1})}{\Phi[(2y_{it2}-1)w_{it2}\pi_2]}x_{it1} \\
&\equiv \frac{y_{it1}-m_{it}}{m_{it}(1-m_{it})}g_{it1} \\
v_{it1} &= \frac{(2y_{it2}-1)w_{it2}\pi_2 - (2y_{it2}-1)\rho_t w_{it1}\pi_1}{\sqrt{1-\rho_t^2}},
\end{aligned} \tag{2.15}$$

where $\phi(\cdot)$ indicates the standard normal probability density function (pdf) and I define g_{it1} to simplify notations for the Hessian matrix.

In the same way,

$$\frac{\partial \ln \ell_{it}}{\partial \pi_2} = \frac{y_{it1} - m_{it}}{m_{it}(1 - m_{it})} \frac{(2y_{it2} - 1)\phi(w_{it2}\pi_2)\phi(v_{it2})}{\Phi[(2y_{it2} - 1)w_{it2}\pi_2]} x_{it2} \tag{2.16}$$

$$+ \frac{y_{it2} - \Phi(w_{it2}\pi_2)}{\Phi(w_{it2}\pi_2)\left[1 - \Phi(w_{it2}\pi_2)\right]} \left[y_{it2} - \Phi(w_{it2}\pi_2)\right] x_{it2}$$

$$= \frac{y_{it1} - m_{it}}{m_{it}(1 - m_{it})} g_{it2} + \frac{y_{it2} - \Phi(w_{it2}\pi_2)}{\Phi(w_{it2}\pi_2)\left[1 - \Phi(w_{it2}\pi_2)\right]} u_{it2} x_{it2}$$

$$v_{it2} = \frac{w_{it1}\pi_1 - (2y_{it2} - 1)\rho_t(2y_{it2} - 1)w_{it2}\pi_2}{\sqrt{1 - \rho_t^2}}.$$

Because π_2 appears in both the structural equation and the reduced form for y_2, there is an additional term.

The score function for ρ_t is

$$\frac{\partial \ln \ell_{it}}{\partial \rho_t} = \frac{y_{it1} - m_{it}}{m_{it}(1 - m_{it})} \tag{2.17}$$

$$\times \frac{(2y_{it2} - 1)\phi_2[w_{it1}\pi_{1t}, (2y_{it2} - 1)w_{it2}\pi_2, (2y_{it2} - 1)\rho_t]}{\Phi[(2y_{it2} - 1)w_{it2}\pi_2]}$$

$$\equiv \frac{y_{it1} - m_{it}}{m_{it}(1 - m_{it})} g_{it\rho},$$

where $\phi_2(\cdot)$ denotes the bivariate standard normal pdf.

Some components of $\nabla_\pi^2 \ln \ell_{it}(\pi)$ are

$$\frac{\partial^2 \ln \ell}{\partial \pi_1 \partial \pi_1'} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{-1}{m_{it}(1 - m_{it})} g_{it1}^2 x_{it1} x_{it1}', \tag{2.18}$$

$$\frac{\partial^2 \ln \ell}{\partial \pi_1 \partial \rho_t'} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{-1}{m_{it}(1 - m_{it})} g_{it1} g_{it\rho} x_{it1},$$

$$\frac{\partial^2 \ln \ell}{\partial \pi_1 \partial \pi_2'} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{-1}{m_{it}(1 - m_{it})} g_{it1} g_{it2} x_{it1} x_{it2}'.$$

The remaining elements of the Hessian matrix can be obtained in the similar way.

The test for time-varying individual heterogeneity is conducted based on the more general function for y_1 as in the first chapter. See the first chapter for detailed explanation about the more general function. Also the LM (score) test, the variable addition test, and the minimum distance test are conducted as before.

2.4 Application: fertility effect on working hours of women

I apply the method to study the effect of fertility on working hours of women with two or more children. As I mentioned above, the same sex indicator is used as an instrumental variable. Therefore, the interpretation of the average treatment effects of a new-born is the effect of fertility moving from the second child to the third one on fractions of working hours of women with two or more children.

Panel Study of Income Dynamics (PSID) for 1986-1989 is used as in Carrasco (2001). I refer readers to Carrasco (2001) for detailed explanation about the data. The PSID is longitudinal survey over 5000 households since 1968 and consists of two independent subsamples: equal probability sample of US household and below 150% of the poverty line. The data in this paper includes both subsamples to increase sample size. The data contains 1442 married or cohabitating women with two or more children between the ages of 18 and 55 in 1986. Table (B.1) shows descriptive statistics for key variables. The women spent about 19% of their total hours to work and there is only one observation of 1. 14.3% of women had a new born baby and 37% of them had children aged between 2 and 6 years in 1986. Their average age is 31.302 and about 77.3% of them are black. As the ratio of women have new born babies decreases, the fraction of working hours to total hours of women increases.

The conditional mean function for labor supply is

$$E(fwh_{it}|\cdot) = \Phi[\frac{\psi_t T_t + \alpha fert_{it} + kid26_{it}\beta_1 + hinc_{it}\beta_2 + black_i\beta_3}{\sqrt{1+\tau_t^2}} + \frac{age_{it}\beta_4 + \eta_t\overline{dsex}_i\zeta_1 + \eta_t\overline{kid26}_i\zeta_2 + \eta_t\overline{hinc}_i\zeta_3 + \rho_t r_{it2}}{\sqrt{1+\tau_t^2}}], \quad (2.19)$$

where T_t indicates a set of year dummies. fwh is the ratio of working hours to total available hours and $fert$ is a fertility variable which equals 1 if the mother has a child aged 1 at $t+1$, because it is not clear that a child who is a year old at the time of interview means a mother gives birth to the child at that year or previous year [Carrasco (2001)]. Two more exogenous variables are added: $kid26$ takes 1 if a woman has children aged between 2 and 6 years and $hinc$ is a logarithm form of husband's income. Mothers' race ($black$) and age (age) are also controlled. The time averages of exogenous variables are added to control individual heterogeneity. $dsex$ is a same sex indicator, which equals 1 if the first two children have the same sex. This variable is used as an instrument variable for the fertility based on parental preference on mixed siblings. Because of indistinguishability and exogeneity of $black$ and age, their time averages do not appear in the conditional mean function for fwh. To consider the endogeneity of $fert$, a linear projection of $r_{it1} = \rho_t r_{it2} + \varepsilon_{it}$, $\varepsilon_{it}|z_i \sim N(0, \tau_t^2.)$ is assumed. r_{it1} denotes an unobserved omitted variable in the model for fwh and is assumed to be correlated with $fert$. r_{it2} appears in the reduced form for $fert$.

Equation (2.20) shows the reduced form for $fert$.

$$
\begin{aligned}
E(fert_{it}|\cdot) &= \Phi(\omega_t T_t + \delta dsex_{it} + kid26_{it}\gamma_1 + hinc_{it}\gamma_2 + black_i\gamma_3 \\
&\quad + age_{it}\gamma_4 + \overline{dsex}_i\lambda_1 + \overline{kid26}_i\lambda_2 + \overline{hinc}_i\lambda_2 + r_{it2}).
\end{aligned} \tag{2.20}
$$

To obtain the conditional mean function for fwh, the bivariate probit model is used and the result looks like equation (2.11). The conditional mean function for fwh along with the reduced form for $fert$ is estimated by Bernoulli QMLE.

For the purpose of comparison, a linear model is also estimated and reported. Let us explain how to estimate the linear model before looking at estimation results. In strictly exogenous case, the fixed effects estimator is used. In binary EEVs case, I use the two-step estimator.

$$
\begin{aligned}
y_{it1} &= w_{it1}\pi_1 + v_{it1} \\
y_{it2} &= 1(w_{it2}\pi_2 + v_{it2} > 0),
\end{aligned}
$$

where y_{it1} and y_{it2} denote the fractional dependent variable and a binary EEV, respectively. Because the endogeneity of y_{it2} implies $v_{it1} = \rho v_{it2} + e_{it1}$. I define a new regressor, $y_{it2}\phi(w_{it2}\pi_2)/\Phi(w_{it2}\pi_2) - (1 - y_{it2})\phi(w_{it2}\pi_2)/\left[1 - \Phi(w_{it2}\pi_2)\right]$, add it to control the endogeneity. The motivation of this approach is as follows. The expectation of v_{it1} conditional on w_{it1} is $\rho E(v_{it2} \mid w_{it1}) + E(e_{it1} \mid w_{it1})$, where $E(e_{it1} \mid w_{it1}) = 0$ by independency of e_{it1} on w_{it1}. $E(v_{it2} \mid w_{it1})$ can be obtained from $y_{it2}E(v_{it2} \mid v_{it2} > -w_{it2}\pi_2) + (1 - y_{it2})E(v_{it2} \mid v_{it2} < -w_{it2}\pi_2)$. If we assume the probit model for y_{it2}, we can prove $E(v_{it1} \mid w_{it1}) = y_{it2}\phi(w_{it2}\pi_2)/\Phi(w_{it2}\pi_2) - (1 - y_{it2})\phi(w_{it2}\pi_2)/\left[1 - \Phi(w_{it2}\pi_2)\right]$. Therefore, we estimate the probit model for y_{it2} at the first step, and the fixed effects estimator is used after adding the generated regressor [Freedman and Sekhon (2008)]. As Wooldridge pointed out, the fixed effects instrumental variable (FEIV) estimator can be used under weaker assumptions in that the FEIV does not require y_2 follows a probit. Because the linear model is used to compare the coefficients of the linear model with the ATEs of the fractional model and the fractional model assumes the probit reduced form for y_2, I use this Heckman approach.

One can use the plug-in method, where the fitted values for y_{it2} at the first step, $\widehat{\Phi} \equiv \Phi(w_{it2}\widehat{\pi}_2)$, is used an IV for y_{it2}. Because $E(y_{it2} \mid w_{it2}) = \Phi(w_{it2}\pi_2)$ holds and we have a linear model, this plug-in method also produces consistent estimates[4].

Table (B.2) and Table (B.3) show the estimates of various models for strictly exogenous fertility case. Table (B.2) reports the results of the linear and the fractional model. All three methods show significantly negative effect of fertility on fractions of working hours. The coefficient of fertility in the linear model is -0.021 and the ATEs of fertility in the fractional model are -0.023 in both the FPE and MWNLS estimator. Compared to women without new born babies, mothers with the new born reduce their fractions of working hours by 2.3% points. Also table (B.2) contains two test statistics for time-varying heterogeneity. While the LM test statistics accept the null hypothesis of constant individual heterogeneity ($H_0 : \eta_t = 1$), the variable addition test rejects the hypothesis.

Table (B.3) contains the results of the fraction model with time-varying individual heterogeneity. The first column and the third column are the estimates of the constrained

[4] According to the simulation results in this chapter, both two-step estimator and the plug-in estimator show quite similar results in term of the root mean squared erros, compared to the average treatment effects in the population.

pooled FPE and the constrained MWNLS where the restriction of $\pi = h(\theta)$ is imposed directly on the model. The second column contains the estimates of the CMD, where the set of parameters of interest θ is recovered from π by the classical minimum distance estimator. The ATEs are -0.022 ~-0.023 and these number are quite similar with those of the linear and the time-constant heterogeneity model. Also the GMM test statistic cannot reject the null hypothesis of $\eta_t = 1$.

Table (B.4) and table (B.5) contain the results of the linear, the fractional model, and the fractional time-varying model for endogenous fertility case. According to table (B.4), the effect of fertility on the fraction of working hours is -0.142 in the linear model, while the ATEs of fertility in the fractional model is -0.129. Also table (B.5) shows quite similar results with the ATEs of the fractional model in table (B.4). The ATEs in table (B.5) are -0.122 ~ -0.126 depending on estimation methods. In the same way, we can interpret these results as follows: mothers with new born babies reduce their fractions of working hours by 12.9 ~ 12.2 % points compared to women without. According to table (B.1), the mean values of fwh are around 0.2. Therefore, the ATEs of fertility seem not only statistically but also economically significant and meaningful. The exogeneity of fertility is rejected in both the fractional model (the estimate of the correlation between r_1 and r_2 is 0.214 and its standard error is 0.058) and the fractional time-varying model (the p-value of the Wald test for endogeneity is 0.0293), but the time-constant individual heterogeneity cannot be rejected.

2.5 Monte-Carlo simulations

Simulations are conducted to find finite sample properties of a linear model, inconsistent two-step procedure, and the single-step estimator. For the purpose of comparison, the

ATEs of a binary EEV are computed and their mean squared errors and biases are reported. The number of replication is 500 and I generate eight samples with N=300, 500, 700, 1000 and T=5, 7.

The equation (2.21) shows the data generating process for a binary EEV.

$$\begin{aligned} x_{it} &= 1[\gamma_{xt}T_t + \gamma_{xz}z_{it} + \gamma_{xc}c_i + u_{it} > 0], \\ c_i, u_{it} &\sim N(0,1), \end{aligned} \tag{2.21}$$

where x_{it} and z_{it} denote a binary EEV and its instrumental variable, respectively. c_i indicates unobserved individual heterogeneity and is assumed to be correlated with x_{it}. Also T_t (time dummies) are added to the model to allow different intercepts for each time.

$$z_{it} = \gamma_{zc}c_i + \sqrt{1-\gamma_{zc}^2}w_{it}, w_{it} \sim N(0,1). \tag{2.22}$$

Equation (2.22) is the data generating process for z_{it}. Also z_{it} is set to be correlated with c_i.

$$\begin{aligned} p &= \Phi(\gamma_{yt}T_t + \gamma_{yx}x_{it} + c_i + v_{it}), \\ v_{it} &= \rho u_{it} + e_{it}, e_{it} \sim N(0,1), \\ y_{it} &= \frac{1}{H}\sum_{h=1}^{H} y_{ith}, y_{ith} \sim Bernoulli(p), \end{aligned} \tag{2.23}$$

where v_{it} denotes an unobserved omitted variable and is correlated with x_{it} through u_{it}. The fractional dependent variable, y_{it}, is generated based on the Bernoulli distribution with probability of p. As before, I use $H = 1000$ Bernoulli draws. The average treatment effect is approximated as

$$ATE_{t,pop} \approx \frac{1}{N}\sum_{i=1}^{N}\left[\Phi(\gamma_{yt}T_t + \gamma_{yx} + c_i + v_{it}) - \Phi(\gamma_{yt}T_t + c_i + v_{it})\right]. \quad (2.24)$$

The equation (2.24) shows the ATEs for time t in population.

As mentioned above, the root mean squared errors (RMSEs) and biases are compared based on the ATEs of three estimation methods.

1. The fixed effects estimator of a linear model

$$\begin{aligned} y_{it} &= \gamma_{yt}T_t + \gamma_{yx}x_{it} + \delta_{yz}\bar{z}_i + \rho v_{it2}. \quad (2.25) \\ P(x_{it} &= 1|z_i) = \Phi(\alpha_{xt}T_t + \alpha_{xz}z_{it} + \beta_{xz}\bar{z}_i). \end{aligned}$$

(1) Estimate the second equation of (2.25) by the probit estimator and generate a new regressor, $\widehat{v}_{it2} \equiv x_{it}\widehat{\phi}_{it}/\hat{\Phi}_{it} - (1 - x_{it})\widehat{\phi}_{it}/(1 - \hat{\Phi}_{it})$, where $\widehat{\phi_{it}} = \phi(\hat{\alpha}_{xt}T_t + \hat{\alpha}_{xz}z_{it} + \hat{\beta}_{xz}\bar{z}_i)$ and $\hat{\Phi}_{it} = \Phi(\hat{\alpha}_{xt}T_t + \hat{\alpha}_{xz}z_{it} + \hat{\beta}_{xz}\bar{z}_i)$. (2) Replace v_{it2} with $\widehat{v}_{it2}$ and estimate the model by the fixed effects estimator. (3) get $\widehat{\gamma}_{yx}$.

2. The Two-step fractional probit estimator (FPE)

$$E(y_{it}|x_{it}, z_i) = \Phi(\gamma_{yt}T_t + \gamma_{yx}x_{it} + \delta_{yz}\bar{z}_i + \rho v_{it2}) \quad (2.26)$$

(1) Estimate the same equation of (2.26) by the probit estimator and get $\widehat{v}_{it2}$. (2) Substitute v_{it2} with $\widehat{v}_{it2}$ and estimate the first equation by the fractional probit estimator based on the conditional mean function. (3) Compute the ATEs from (2.27).

$$\begin{aligned} \widehat{ATE}_{t,FPE} &= \frac{1}{N}\sum\nolimits_{i=1}^{N}[\Phi(\hat{\gamma}_{yt}T_t + \hat{\gamma}_{yx} + \hat{\delta}_{yz}\bar{z}_i + \widehat{\rho}\widehat{v}_{it2}) \quad (2.27) \\ &\quad -\Phi(\hat{\gamma}_{yt}T_t + \hat{\delta}_{yz}\bar{z}_i + \widehat{\rho}\widehat{v}_{it2})]. \end{aligned}$$

3. Single-step fractional bivariate probit estimator (FBPE)

$$E(y_{it}|x_{it}, z_i) = \frac{\Phi_2[w_{it1}\pi_1, (2x_{it}-1)w_{it2}\pi_2, (2x_{it}-1)\rho]}{\Phi[(2x_{it}-1)w_{it2}\pi_2]}. \tag{2.28}$$

(1) Estimate the equation (2.28) by the pooled QMLE. (2) Calculate the ATEs from (2.29).

$$\widehat{ATE}_{t,FBPE} = \frac{1}{N}\sum_{i=1}^{N}\left[\Phi\left(\frac{\hat{\gamma}_{yt}T_t + \hat{\gamma}_{yx}x_{it} + \hat{\delta}_{yz}\bar{z}_i}{\sqrt{1+\hat{\rho}^2}}\right) - \Phi\left(\frac{\hat{\gamma}_{yt}T_t + \hat{\delta}_{yz}\bar{z}_i}{\sqrt{1+\hat{\rho}^2}}\right)\right] \tag{2.29}$$

Table (B.6) contains simulation results. The numbers reported are the RMSEs and biases of three estimation methods. As N and T increase, the RMSEs decrease. The RMSEs of the FBPE gives the best fits among three methods. For examples, the RMSE of the FBPE is 0.0238 and the RMSEs of the linear and the FBE are 0.0295 and 0.0294, respectively when $N = 1000$ and $T = 7$. The FPE and FE estimators are inconsistent in this setting, but the FPE performs somewhat better. The nonlinearity in the population seems to lead to this result.

2.6 Conclusion

In this chapter, I extend the first chapter to a binary endogenous explanatory variable (EEV) case. Because the EEV is discrete, the traditional two-stage procedure of Rivers and Vuong (1988) produces inconsistent estimates. Alternatively, the full-information maximum likelihood estimator (FIMLE) can be used. However, if distributional assumption is not true in population or there is misspecification in any equation of the system, the FIMLE also is an inconsistent estimator. I modify the bivariate probit model to derive a conditional mean function for a fractional dependent variable with a binary endogenous variable. Based on this mean function, the pooled quasi-limited information maximum likelihood estimator is proposed. For identification, the conditional mean function along with a reduced form for a binary endogenous variable are estimated [Wooldridge (2007)]. This single-step estimator gives simple asymptotic variances and possibly enhances efficiency of estimates. The estimation methods are applied to the effect of fertility on women's fraction of working hours. Based on the parental preference on the mixed siblings, the same sex indicator of the first two children is used as an instrumental variable to consider the endogeneity of fertility. The ATEs of fertility are -0.023 under exogenous fertility assumption and -0.129 under endogenous fertility assumption. The exogeneity of fertility assumption is rejected, but time-constant individual heterogeneity cannot be rejected. The ATEs of fertility imply that mothers with a new born reduce their working hours by 12.9% points compared to women without new born babies. Also simulations are conducted and the ATEs of a binary endogenous regressor are compared in a linear model, the inconsistent two-step fractional probit estimator (FPE), and the fractional bivariate probit estimator (FBPE). The FBPE gives the best fit among the estimation models and methods.

Chapter 3 THE HURDLE MODEL FOR FRACTIONAL RESPONSE VARIABLES WITH BINARY ENDOGENOUS EXPLANATORY VARIABLES AND TIME-VARYING INDIVIDUAL HETEROGENEITY

3.1 Introduction

This chapter studies the hurdle model (or, the generalized Tobit model) for fractional variables. Also time-varying individual heterogeneity and binary endogenous variables (EEVs) are considered as in the second chapter. The fractional variables can take 0 or 1 with positive probabilities. This stack of 0's or 1's occurs because fractional variables cannot be defined outside the unit interval or are corner solutions (optimal choices). Our two applications in the first and second chapter show two different examples. In the first chapter, nontrivial proportion of school district reports 100% of test pass rates because test pass rates cannot exceed 1. In the second chapter, some mothers report 0 working hours and this stack of 0's seems to be the result of optimal choices.

The fractional hurdle model allows us to separate the determination of corner solutions ($y = 0$ or $y > 0$) from the decision of the amount of y conditional on $y > 0$, where y denotes a fractional dependent variable. By using the fractional hurdle model, we can allow different sets of explanatory variables for the equation for corner solutions and the equation for positive fractional dependent variables. Also, we can obtain both $E(y \mid x)$ and $E(y \mid x, y > 0)$, and their corresponding average treatment effects (ATEs) can be calculated. The motivation of the second chapter is that the effect of fertility might be underestimated when the binary model for job market participation is used because the binary model ignores women who are working but reduce their working hours after having new babies. The ATEs of fertility in $E(y \mid x, y > 0)$ shows how much the binary model possibly underestimate the effect of new born babies.

After deriving a conditional mean function, the single step estimator of Wooldridge (2007) is used. To obtain a conditional mean function the bivariate probit model is modified as before. The estimation methods are applied to analyze the fertility effect on women's labor supply by using the same data as in the second chapter. Two average treatment effects (ATEs) of the fertility are obtained and compared with those in chapter 2. The first ATEs are obtained using the whole sample and the second ATEs are calculated conditional on the fractions of working hours are greater than 0.

3.2 The basic fractional hurdle model

I assume N randomly drawn cross-sectional units and T observations for each cross-sectional unit. Also N is assumed to be greater than T (large N fixed T asymptotics). The basic hurdle model for fractional response variables consists of two equations. The first equation is (3.1), which indicates the probability that the fractional dependent variable, y_{it}, takes 0 or positive values.

$$\begin{aligned} p(d_{it} &= 1|x_{it}, c_{i1}) = p(y_{it} > 0|x_{it}, c_{i1}) = \Phi(x_{it}\beta_1 + c_{i1}). \qquad (3.1) \\ c_{i1}|x_i &\sim N(\psi_1 + \bar{x}_i\beta_2 + a_{i1}), var(a_{i1}|x_i) = \sigma_{a1}^2, \end{aligned}$$

where i indicate a cross sectional unit and t indicate time ($i = 1, 2, ..., N, t = 1, 2, ...T$). d_{it} equals 1 when y_{it1} takes positive values. x_{it} and c_{i1} denote a set of explanatory variables and individual heterogeneity, respectively. The individual heterogeneity is assumed invariant over time. One can easily consider time-varying heterogeneity by using interactive effects model. The distribution of c_1 is restricted by using the Mundlak(1978) - Chamberlain(1980) device.

From the properties of the normal distribution,

$$\begin{aligned} p(d_{it} &= 1|x_i) = \Phi\left(\psi_1 + x_{it}\beta_1 + \bar{x}_i\beta_2 + a_{i1}\right) \qquad (3.2)\\ &= \Phi\left(\frac{\psi_1 + x_{it}\beta_1 + \bar{x}_i\beta_2}{\sqrt{1+\sigma_{a1}^2}}\right) \\ &\equiv \Phi(w_{it}\pi_1), w_{it} \equiv (1, x_{it}, \bar{x}_i), \end{aligned}$$

where $x_i \equiv (x_{i1}, ..., x_{iT})$ is a set of explanatory variables in all times and $\bar{x} \equiv N^{-1}\sum_{i=1}^{N} x_{it}$ is a set of time averages of x_{it}. Because only scaled parameters are identified, I drop subscript a from now on and define w_{it} and π_1 for notational simplicity.

The second equation is (3.3), which shows the mean function conditional on $y > 0$.

$$\begin{aligned} E(y_{it}|x_{it}, c_{i2}, y_{it} &> 0) = \Phi(x_{it}\delta_1 + \eta_t c_{i2}). \qquad (3.3)\\ c_{i2}|x_i &\sim N(\psi_2 + \bar{x}_i\delta_2 + a_{i2}, \sigma_{a2}^2), \end{aligned}$$

where $\eta_t c_{i2}$ is added to consider time varying individual heterogeneity. Applying the Mundlak (1978) - Chamberlain (1980) device leads the equation (3.4).

$$\begin{aligned} E(y_{it}|x_i, y_{it} &> 0) = \Phi\left(\eta_t\psi_2 + x_{it}\delta_1 + \eta_t\bar{x}_i\delta_2 + \eta_t a_{i2}\right) \qquad (3.4)\\ &= \Phi\left(\frac{\eta_t\psi_2 + x_{it}\delta_1 + \eta_t\bar{x}_i\delta_2}{\sqrt{1+\eta_t^2\sigma_{a2}^2}}\right) \equiv \Phi(w_{it}\pi_{2t}). \end{aligned}$$

Therefore, the conditional mean function for y is

$$E(y_{it}|x_i) = p(y_{it}=0|x_i)\times 0 + p(y_{it}>0|x_i)\times E(y_{it}|x_i, y_{it}>0) \tag{3.5}$$
$$= \Phi(w_{it}\pi_1)\Phi(w_{it}\pi_{2t}) = \Phi_2(w_{it}\pi_1, w_{it}\pi_{2t}, 0) \equiv m(w_{it}, \pi_t),$$

where Φ_2 indicates the bivariate standard normal cumulative density function (cdf). The conditional mean function turns out to be the product of two standard normal cdfs and we can rewrite the mean function by using the bivariate standard normal cdf with the correlation of 0. I will use this property later to obtain the average treatment effects (ATEs) of a binary explanatory variable.

The partial effects of the kth explanatory variable, x_k, can be obtained either from taking derivatives of the conditional mean function with respect to x_k if x_k is continuous, or from the differences of the mean function evaluated at two values of x_k if x_k is discrete. As before, these partial effects depend not only observables but unobserved c_1 and c_2. Therefore, it is a good idea to calculate the average partial effects by averaging partial effects across the distributions of c_1 and c_2.

$$\begin{aligned} APE_t &= E_{(x,c_1,c_2)}[\beta_k\phi(x_t\beta_1 + c_{i1})\Phi(x_t\delta_1 + \eta_t c_{i2}) \\ &\quad + \beta_k\Phi(x_t\beta_1 + c_{i1})\phi(x_t\delta_1 + \eta_t c_{i2})] \\ ATE_t &= E_{(x,c_1,c_2)}[\Phi(x_t^{(1)}\beta_1 + c_{i1})\Phi(x_t^{(1)}\delta_1 + \eta_t c_{i2}) \\ &\quad - \Phi(x_t^{(0)}\beta_1 + c_{i1})\Phi(x_t^{(0)}\delta_1 + \eta_t c_{i2})], \end{aligned} \tag{3.6}$$

where APE_t and ATE_t stands for the average partial effects and the average treatment effects of x_k at t, respectively and $x_{it}^{(1)}$ and $x_{it}^{(0)}$ denote sets of explanatory variables for two different values of x_{itk}. Exploiting the Mundlak (1978) - Chamberlain (1980) device and assuming the independency of (a_{i1}, a_{i2}) conditional on x_i, we can rewrite (3.6) as

$$\begin{aligned} APE_t &= E_{\bar{x}_i}\left(\frac{\beta_k}{\sqrt{1+\sigma_{a1}^2}}\right)\phi(w_t\pi_1)\Phi(w_t\pi_{2t}) \qquad (3.7) \\ &+\left(\frac{\beta_k}{\sqrt{1+\eta_t^2\sigma_{a2}^2}}\right)\Phi(w_t\pi_1)\phi(w_t\pi_{2t}) \\ ATE_t &= E_{\bar{x}_i}\left[\Phi(w_t^{(1)}\pi_1)\Phi(w_t^{(1)}\pi_{2t}) - \Phi(w_t^{(0)}\pi_1)\Phi(w_t^{(0)}\pi_{2t})\right], \end{aligned}$$

where $w_{it}^{(1)}$ and $w_{it}^{(0)}$ denote sets of explanatory variables for two different values of w_{it}.

3.3 Estimation methods under strict exogeneity

Based on the equation (3.5), we can always use the pooled nonlinear least squares (NLS) estimator. This pooled NLS estimator gives consistent estimates, but might be inefficient because of possible heteroskedasticity and/or serial correlation in errors. In this paper, the pooled quasi-maximum likelihood estimator (QMLE) and the multivariate nonlinear least squares (MWNLS) estimator to consider the serial correlation in errors are used. Also the conditional mean function for y along with the probit model for d is estimated because of the identification problem. The log likelihood function for the pooled Bernoulli QMLE is

$$\begin{aligned} l_{it} &= y_{it}\ln m(w_{it},\pi_t) + (1-y_{it})\ln[1-m(w_{it},\pi_t)] \qquad (3.8) \\ &+d_{it}\ln\Phi(w_{it}\pi_1) + (1-d_{it})\ln[1-\Phi(w_{it}\pi_1)]. \end{aligned}$$

The MWNLS estimator might enhance the efficiency by using the possible correlation in errors within a cross-sectional unit. The MWNLS estimator solves

$$\min_{\theta} \sum_{i=1}^{N} [d_i - \Phi(w_i\pi_1)]' \widetilde{V}_{i1}^{-1} [d_i - \Phi(w_i\pi_1)] \qquad (3.9)$$
$$+ [y_i - m(w_i, \pi)]' \widetilde{V}_{i2}^{-1} [y_i - m(w_i, \pi)]$$
$$\widetilde{V}_{i1} = Diag\left[\widetilde{\Phi}_i(1-\widetilde{\Phi}_i)\right]^{1/2} C(\widetilde{\rho_2}) Diag\left[\widetilde{\Phi}_i(1-\widetilde{\Phi}_i)\right]^{1/2}$$
$$\widetilde{V}_{i2} = Diag\left[\widetilde{m_i}(1-\widetilde{m_i})\right]^{1/2} C(\widetilde{\rho_1}) Diag\left[\widetilde{m_i}(1-\widetilde{m_i})\right]^{1/2},$$

where $m_i \equiv m(w_i, \pi)$ and $\Phi_i \equiv \Phi(w_i\pi_1)$. '$\sim$' indicates that all elements are evaluated at the preliminary estimates. After estimating (3.2) and (3.5) by the nonlinear least squares (NLS) estimator, one can calculate the correlations ($\widetilde{\rho_1}$, $\widetilde{\rho_2}$) by using the standardized residuals from the NLS estimates. See Papke and Wooldridge (2008) and the first chapter for detailed explanation.

With consistent estimates, the ATEs can be obtained from (3.10).

$$\widehat{ATE}_t = \frac{1}{N}\sum_{i=1}^{N}\left[\Phi(w_t^{(1)}\hat{\pi}_1)\Phi(w_t^{(1)}\hat{\pi}_{2t}) - \Phi(w_t^{(0)}\hat{\pi}_1)\Phi(w_t^{(0)}\hat{\pi}_{2t})\right]. \qquad (3.10)$$

3.4 Estimation methods with binary endogenous explanatory variables

Now relax the strict exogeneity assumption. I consider a binary endogenous explanatory variable (EEV) case with appropriate instrument variables. Estimation methods with continuous EEVs can be found in the first chapter. In continuous EEVs case, the two-step estimator by Rivers and Voung (1988) can be used and this estimator produces consistent estimates. However, there is no convenient two-step estimators but inconsistent plug-in estimator with binary EEVs (Wooldridge, 2007).

The equation for corner solutions is defined as follows.

$$
\begin{aligned}
d_{it} &= 1(\alpha_1 y_{it2} + z_{it1}\beta_1 + c_{i1} + v_{it1} > 0), d_{it} \equiv (y_{it1} > 0). \qquad (3.11)\\
c_{i1}|z_i &\sim N(\psi_1 + \bar{z}_i\beta_2 + a_{i1}, \sigma_{a1}^2),
\end{aligned}
$$

where $\bar{z}_i \equiv 1/T\sum_{t=1}^{T} z_{it}$, $z_{it} = (z_{it1}, z_{it2})$. z_{it} contains both explanatory variables (z_{it1}) and instrumental variables (z_{it2}). y_{it2} is a binary EEV and v_{it1} denotes an unobserved omitted variable that is assumed to be correlated with y_{it2}. The general index function, $1(\cdot)$, is used, but I assume the probit model for d to exploit the properties of the standard normal cumulative density function (cdf).

From the Mundlak (1978) - Chamberlain (1980) device,

$$
\begin{aligned}
d_{it} &= 1\left(\psi_1 + \alpha_1 y_{it2} + z_{it1}\beta_1 + \bar{z}_i\beta_2 + a_{i1} + v_{it1} > 0\right) \qquad (3.12)\\
&= 1\left(\psi_1 + \alpha_1 y_{it2} + z_{it1}\beta_1 + \bar{z}_i\beta_2 + r_{it1} > 0\right)\\
&\equiv 1(w_{it1}\pi_1 + r_{it1} > 0), w_{it1} \equiv (1, z_{it1}, \bar{z}_i).
\end{aligned}
$$

The following equation shows the mean function conditional on $y > 0$.

$$
\begin{aligned}
E(y_{it1}|y_{it2}, z_{it1}, c_{i2}, v_{it2}, y_{it1} &> 0) = \Phi(\alpha_2 y_{it2} + z_{it1}\delta_1 + \eta_t c_{i2} + v_{it2}) \qquad (3.13)\\
c_{i2}|x_i &\sim N(\psi_2 + \bar{z}_i\delta_2 + a_{i2}, \sigma_{a2}^2),
\end{aligned}
$$

where v_{it2} denotes an unobserved omitted variable and is assumed to be correlated with y_{it1}.

In the same way, we can rewrite the mean function as

$$
\begin{aligned}
E(y_{it1}|y_{it2}, z_i, r_{it2}, y_{it1} &> 0) = \Phi[\eta_t\psi_2 + \alpha_2 y_{it2} + z_{it1}\delta_1 \\
&\quad + \eta_t \bar{z}_i \delta_2 + r_{it2}] \\
&\equiv \Phi(w_{it1}\pi_2 + r_{it2}), r_{it2} \equiv \eta_t a_{i2} + v_{it2}.
\end{aligned} \tag{3.14}
$$

Now let's define the reduced form for y_2. Because the reduced form is also a nonlinear function, the Mundlak(1978)-Chamberlain(1980) device is applied.

$$
\begin{aligned}
y_{it2} &= 1(z_{it}\gamma + c_{i3} + v_{it3} > 0). \\
c_{i3}|z_{it} &\backsim N(\psi_3 + \bar{z}_i\zeta + a_{i3}, \sigma_{a3}^2),
\end{aligned} \tag{3.15}
$$

We can rewrite (3.15) as

$$
\begin{aligned}
y_{it2} &= 1(\psi_3 + z_{it}\gamma + \bar{z}_i\zeta + r_{it3} > 0) \\
&\equiv 1(w_{it2}\pi_3 + r_{it3} > 0), r_{it3} \equiv a_{i3} + v_{it3}.
\end{aligned} \tag{3.16}
$$

The endogeneity of y_{it2} implies that r_{it3} is correlated with both r_{it1} but r_{it2}. Therefore, linear projections of r_{it1} and r_{it2} on r_{it3} are assumed.

$$
\begin{aligned}
r_{it1} &= \rho_1 r_{it3} + e_{it1}, V(e_{it1}|z_{it}) = \tau_1^2. \\
r_{it2} &= \rho_{2t} r_{it3} + e_{it2}, V(e_{it2}|z_{it}) = \tau_{2t}^2.
\end{aligned} \tag{3.17}
$$

After assuming joint normality $(r_{it1}, r_{it2}, r_{it3})$, I modify the multivariate probit model to obtain the conditional mean function for y_{it1} by taking expectation.

Equation (3.12) can be rewritten as follows.

$$
\begin{aligned}
p(d_{it} &= 1|y_{it2}, z_i, r_{it1}) = \Phi(w_{it1}\pi_1 + r_{it1}) \qquad (3.18)\\
p(d_{it} &= 1|y_{it2}, z_i, r_{it3}) = \Phi\left(\frac{w_{it1}\pi_1 + \rho_1 r_{it3}}{\sqrt{1+\tau_1^2}}\right)\\
&\equiv \Phi(w_{it1}\pi_1 + \rho_1 r_{it3}).
\end{aligned}
$$

Because only scaled parameters are identified, I drop subscript for simplicity.

From the results of the second chapter,

$$
\begin{aligned}
E_{r3}(d_{it}|y_{it2,}z_i) &= E_{r3}[\Phi(w_{it1}\pi_1 + \rho_1 r_{it3})] \qquad (3.19)\\
&= \frac{y_{it2}}{\Phi(w_{it2}\pi_3)} \int_{-\infty}^{w_{it2}\pi_3} \Phi(w_{it1}\pi_1 + \rho_1 r_{it3})\phi(r_3)dr_3\\
&\quad + \frac{1-y_{it2}}{\Phi(-w_{it2}\pi_3)} \int_{-\infty}^{-w_{it2}\pi_3} \Phi(w_{it1}\pi_1 + \rho_1 r_{it3})\phi(r_3)dr_3\\
&= \frac{\Phi_2[w_{it1}\pi_1, (2y_{it2}-1)w_{it2}\pi_3, (2y_{it2}-1)\rho_1]}{\Phi[(2y_{it2}-1)w_{it2}\pi_2]}\\
&\equiv m_{21it}(w_{it1}, w_{it2}, \pi)
\end{aligned}
$$

In the same way, we can rewrite the equation (3.14) as

$$
\begin{aligned}
E(y_{it1}|y_{it2}, z_{it}, r_{it2}, y_{it1} &> 0) = \Phi(w_{it1}\pi_1 + r_{it1}) \qquad (3.20)\\
&= \Phi\left(\frac{w_{it1}\pi_1 + \rho_{2t} r_{it3}}{\sqrt{1+\tau_{1t}^2}}\right).
\end{aligned}
$$

Taking expectation with respect to r_3 leads

$$
\begin{aligned}
E_{r3}(y_{it1}|y_{it2}, z_i, y_{it1} &> 0) = E_{r3}[\Phi(w_{it1}\pi_{2t} + \rho_{2t}r_{it3})] \qquad (3.21) \\
&= \frac{y_{it2}}{\Phi(w_{it2}\pi_3)} \int_{-\infty}^{w_{it2}\pi_3} \Phi(w_{it1}\pi_{2t} + \rho_{2t}r_{it3})\phi(r_3)dr_3 \\
&\quad + \frac{1 - y_{it2}}{\Phi(-w_{it2}\pi_3)} \int_{-\infty}^{-w_{it2}\pi_3} \Phi(w_{it1}\pi_{2t} + \rho_{2t}r_{it3})\phi(r_3)dr_3 \\
&= \frac{\Phi_2[w_{it1}\pi_{2t}, (2y_{it2} - 1)w_{it2}\pi_3, (2y_{it2} - 1)\rho_{2t}]}{\Phi[(2y_{it2} - 1)w_{it2}\pi_2]} \\
&\equiv m_{22it}(w_{it1}, w_{it2}, \pi).
\end{aligned}
$$

To obtain the conditional mean function for y_1,

$$
\begin{aligned}
E(y_{it1}|y_{it2}, z_i) &= p(d_{it} = 1|y_{it2}, z_i) \times E(y_{it}|y_{it2}, z_i, y_{it1} > 0) \qquad (3.22) \\
&\equiv \frac{\Phi_2[w_{it1}\pi_1, (2y_{it2} - 1)w_{it2}\pi_3, (2y_{it2} - 1)\rho_1]}{\Phi[(2y_{it2} - 1)w_{it2}\pi_2]} \\
&\quad \times \frac{\Phi_2[w_{it1}\pi_{2t}, (2y_{it2} - 1)w_{it2}\pi_3, (2y_{it2} - 1)\rho_{2t}]}{\Phi[(2y_{it2} - 1)w_{it2}\pi_2]} \\
&\equiv m_{21it}(w_{it1}, w_{it2}, \pi) \times m_{22it}(w_{it1}, w_{it2}, \pi) \\
&\equiv m_{2it}(w_{it1}, w_{it2}, \pi),
\end{aligned}
$$

where $\pi \equiv (\pi_{1t}', \pi_{2t}', \pi_3')'$. After estimating the model, we can test the endogeneity of y_2 based on ρ_1 and ρ_{2t}. To identify all parameters and possibly enhance the efficiency of estimates, the equation (3.22) is estimated along with (3.19) and (3.16) by the pooled QMLE.

$$
\begin{aligned}
\ln \ell_{it} &= d_{it}\ln(m_{21it}) + (1-d_{it})\ln(1-m_{21it}) \\
&\quad + y_{it1}\ln(m_{2it}) + (1-y_{it1})\ln(1-m_{2it}) \\
&\quad + y_{it2}\ln[\Phi_2(w_{it2}\pi_3)] + (1-y_{it2})\ln[1-\Phi(w_{it2}\pi_3)].
\end{aligned} \tag{3.23}
$$

Now let us derive the ATEs of the binary EEV, y_2. Two assumptions hold: y_2 is independent of (c_1, v_1, c_2, v_2) conditional on $(z, r3)$ and $(z, r3)$ are redundant in the structural mean function. From Wooldridge (2002 section 2.2.5), the ATEs can be obtained from differences in the following function evaluated at $y_{it2} = 0$ and $y_{it2} = 1$. Dropping the subscript i,

$$
\begin{aligned}
& E_{(c_1,v_1,c_2,v_2)}[\Phi(\alpha_1 y_{t2} + z_{t1}\beta_1 + c_1 + v_{t1}) \\
& \qquad \times \Phi(\alpha_2 y_{t2} + z_{t1}\delta_1 + \eta_t c_2 + v_{t2})] \\
&= E_{(z,r_1,r_2)}[\Phi(w_{t1}\pi_1 + r_{t1})\Phi(w_{t1}\pi_{2t} + r_{t2})] \\
&= E_{(z,r3)}[\Phi(w_{t1}\pi_1 + \rho_1 r_{t3})\Phi(w_{t1}\pi_{2t} + \rho_{2t} r_{t3})]
\end{aligned} \tag{3.24}
$$

By exploiting the properties of normal cdfs after transforming the errors, we can rewrite (3.24) as follows and consistent estimator for the ATEs is

$$
\begin{aligned}
\widehat{ATE}_t &= \frac{1}{N}\sum_{i=1}^{N}[\Phi_2\left(\frac{w_{t1}^{(1)}\hat{\pi}_1}{\sqrt{1+\hat{\rho}_1^2}}, \frac{w_{t1}^{(1)}\hat{\pi}_{2t}}{\sqrt{1+\hat{\rho}_{2t}^2}}, \frac{\hat{\rho}_1\hat{\rho}_{2t}}{\sqrt{1+\hat{\rho}_1^2}\sqrt{1+\hat{\rho}_{2t}^2}}\right) \\
&\quad -\Phi_2\left(\frac{w_{t1}^{(0)}\hat{\pi}_1}{\sqrt{1+\hat{\rho}_1^2}}, \frac{w_{t1}^{(0)}\hat{\pi}_{2t}}{\sqrt{1+\hat{\rho}_{2t}^2}}, \frac{\hat{\rho}_1\hat{\rho}_{2t}}{\sqrt{1+\hat{\rho}_1^2}\sqrt{1+\hat{\rho}_{2t}^2}}\right)],
\end{aligned} \tag{3.25}
$$

where $w_{it1}^{(1)}$ and $w_{it1}^{(0)}$ denote sets of explanatory variables with $y_{it2} = 1$ and $y_{it2} = 0$, respectively.

Testing endogeneity of y_2 ($H_0 : \rho_1 = \rho_{2t} = 0$) and time-constant individual heterogeneity ($H_0 : \eta_t = 1$) are conducted in the same way in the chapter 1 and chapter 2.

3.5 Application: fertility effect on working hours of women

I apply the estimation methods to study the effect of fertility on working hours of women with two or more children using the same data as in the second chapter. The main contribution of this chapter is to separate the determination of labor market participation from the determination of fractions of working hours of women. According to the results from the second chapter, the effect of a new born on mother's working hours is significantly negative after controlling the endogeneity of fertility. This implies that women outside job market would not take part in the market after delivering babies. Therefore, focusing on women who remain in the market produces more meaningful implication.

The estimation model for corner solutions is

$$\begin{aligned} E(d_{it} &= 1|\cdot) = \Phi(\psi_{1t}T_t + \beta_1 fert_{it} + kid26_{it}\beta_2 + hinc_{it}\beta_3 \\ &\quad + black_i\beta_4 + age_{it}\beta_5 + \overline{dsex}_i\beta_6 + \overline{kid26}_i\beta_7 + \overline{hinc}_i\beta_8 + r_{it1}), \end{aligned} \tag{3.26}$$

where fwh is the fraction of working hours to total hours and $fert$ is a fertility dummy, which is 1 if a woman has the new born at t+1. Two time-varying explanatory variables are added: $kid26$ is 1 if a woman has children aged between 2 and 6 and $hinc$ is logarithm of husband's income. $dsex$ is 1 if a woman has the same sex children and is used as an instrument variable (IV) for the fertility. This IV is based on parental preference on mixed siblings. Also $black$ and age are controlled. However, because they are likely to be independent on individual heterogeneity and we may not identify the coefficients, their time averages do not appear in the equation. T_t indicates a set of all time dummies to allow different intercepts.

The mean function conditional on positive fractional variable is

$$\begin{aligned} E(fwh_{it}|fwh_{it} &> 0,\cdot) = \Phi(\psi_{2t}T_t + \delta_1 fert_{it} + kid26_{it}\delta_2 + hinc_{it}\delta_3 \\ &+black_i\delta_4 + age_{it}\delta_5 + \overline{dsex}_i\delta_6 + \overline{kid26}_i\delta_7 + \overline{hinc}_i\delta_8 + r_{it2}). \end{aligned} \tag{3.27}$$

In this chapter, I use the same explanatory variables for both the determination equation for a corner solution and the mean function conditional on a positive dependent variable. However, it is straightforward and sometimes desirable to allow different explanatory variables for each equation. In our application, some regressors might affect the decision of whether mothers participate the labor market, but might not affect the amount of hours worked.

The probit model is assumed for fertility.

$$\begin{aligned} E(fert_{it}|\cdot) &= \Phi(\omega_{3t}T_t + \gamma_1 dsex_{it} + kid26_{it}\gamma_2 + hinc_{it}\gamma_3 \\ &+black_i\gamma_4 + age_{it}\gamma_5 + \overline{dsex}_i\gamma_6 + \overline{kid26}_i\gamma_7 + \overline{hinc}_i\gamma_8 + r_{it3}). \end{aligned} \tag{3.28}$$

Table (C.1) contains the estimates of the fractional hurdle model under strictly exogeneity assumption. The first 3 rows reports the estimates of equation (3.26) and the next 3 rows are estimates of equation (3.27). The reported ATEs are from equation (3.10). I report two ATEs: the ATEs conditional on $fwh > 0$ and the ATEs. The ATEs conditional on $fwh > 0$ are obtained based on estimates of equation (3.27). The ATEs of fertility is -0.022 in both the QMLE and the MWNLS estimator and the ATEs conditional on $fwh > 0$ are -0.014~-0.013. The ATEs are quite similar with those of the simple fraction model, but the ATEs conditional on $fwh > 0$ are about half of the ATEs. This difference occurs because some women drop out of the labor market after delivering babies and the ATEs conditional on $fwh > 0$ do not count of them. Both the LM test and the variable addition test accept the hypothesis of constant individual heterogeneity ($H_0 : \eta_t = 1$).

Table (C.2) reports the estimates of the hurdle fractional model with time-varying individual heterogeneity. As we can see, the GMM test statistic also accepts the hypothesis. Therefore, it is not surprising that the ATEs are quite similar with those of the fractional model with constant heterogeneity.

The next two tables (table (C.3) and table (C.4)) contain the results under endogenous fertility assumption. The same sex indicator is used as an instrumental variable. In table (C.3), the first 4 rows report estimates of equation (3.26) and the next 4 rows contain estimates of equation (3.27). In table (C.3) and table (C.4), the ATEs are obtained by using formula (3.25) and the ATEs conditional on $fwh > 0$ are obtained by using the formula in the second chapter based on estimates of equation (3.27). The ATEs are -0.129 and quite similar with those of the simple fractional model. Time-varying individual heterogeneity hypothesis is rejected and the endogeneity of fertility is accepted marginally. In table (C.4), we can interpret the results similarly in table (C.3). The ATEs are -0.124 ~-0.117 depending on the estimation methods. the ATEs conditional on $fwh > 0$ are -0.07 ~-0.09.

Even though the point estimates of η_t are 1.131 ~0.037, their standard errors are too big. The time-varying heterogeneity is rejected by all test statistics. This might imply that we cannot estimate the models precisely because of their nonlinearity and the number of parameters estimated. Therefore, let's focus on the results of the model with time-constant individual heterogeneity. As a result, fertility is marginally endogenous in the model and individual heterogeneity is time-constant. The ATEs of fertility are significantly negative, but the values are quite similar in both the fractional model and the hurdle fractional model. the ATEs conditional on $fwh > 0$ imply that the women who are still working after delivering babies reduce their fraction of working hours by 9 % points. This fertility effect on working women is usually ignored in the binary model, because their binary indicators for labor market participation after having babies do not change.

3.6 Conclusion

In this chapter, I consider the hurdle model for fractional dependent variables. Binary EEVs and time-varying individual heterogeneity are considered as before. With the hurdle model, we can separate the determination of corner solutions from the decision of the amount of dependent variables. In the application, some women report 0 working hour and these might be their optimal choices.

By using the fractional hurdle model, we can allow different sets of explanatory variables for each equation and calculate the average treatment effects (ATEs) conditional on the positive dependent variable. These ATEs are ignored in the binary model.

To derive a conditional mean function, the bivariate probit model is modified. Based on the derived conditional mean function, the pooled QMLE is used and the reduced form for the binary endogenous variable and the mean function for corner solutions are estimated together for identification. I propose how to compute the average treatment effects (ATEs) of binary endogenous variables. The estimation methods are applied to analyze the effect of fertility on women's fraction of worked hours. The ATEs are -0.129 and the ATEs conditional on $fwh > 0$ are -0.09 under endogenous fertility assumption. The endogeneity of fertility is accepted marginally, but the time-varying heterogeneity is rejected by all three test statistics.

Appendix A Tables for Chapter 1

Table A.1: Descriptive Statistics

Year		1995	1996	1997	1998	1999	2000	2001
Test pass rate	Mean	0.618	0.622	0.598	0.748	0.742	0.775	0.756
	Std. dev.	0.131	0.145	0.148	0.129	0.132	0.123	0.126
	Min	0.179	0.152	0.090	0.260	0.250	0.266	0.210
	Max	0.986	1.000	0.970	1.000	1.000	1.000	1.000
avgrexpp	Mean	5514.918	5803.300	6095.992	6326.610	6440.589	6565.288	6704.718
	Std. dev.	984.084	963.712	939.730	922.648	899.656	877.487	876.666
	Min	3966.793	4310.906	4462.045	4831.553	5059.371	5219.352	5280.542
	Max	11497.060	11589.880	11554.280	11348.090	11387.290	11412.310	11444.630
lunch	Mean	0.280	0.277	0.280	0.288	0.292	0.293	0.308
	Std. dev.	0.152	0.155	0.158	0.161	0.165	0.168	0.170
	Min	0.016	0.000	0.015	0.010	0.009	0.012	0.010
	Max	0.844	0.879	0.892	0.913	0.869	0.862	0.901
enrollment	Mean	3066.674	3104.250	3150.670	3157.211	3094.121	3076.950	3068.433
	Std. dev.	8141.043	8270.707	8365.745	8322.376	7764.915	7487.698	7280.209
	Min	192	173	179	171	168	151	172
	Max	170855	173749	175798	174730	161356	154648	149348

Note: *avgrexpp* is the average of the current and previous 3 year spending per pupil in real dollar.

Table A.2: The Linear and the Fractional Model: Exogenous Spending

Model	Linear	Papke and Wooldridge (2008)			
Method	FE	FPE		MWNLS	
Dependent: Test pass rates	Coef	Coef	APE	Coef	APE
log(*avgrexpp*)	0.379	0.883	0.298	0.886	0.298
	(0.071)	(0.206)	(0.07)	(0.206)	(0.07)
lunch	-0.038	-0.206	-0.07	-0.213	-0.072
	(0.073)	(0.207)	(0.069)	(0.207)	(0.068)
log(*enroll*)	0.003	0.092	0.031	0.092	0.031
	(0.049)	(0.138)	(0.045)	(0.138)	(0.045)
LM (score) test		21.652	(0.001)	13.760	(0.032)
Variable addition test		19.160	(0.004)		

Notes: (1) FE, FPE, and MWNSL stand for the fixed effects estimator, the fractional probit estimator, and the multivariate weighted nonlinear least squares estimator, respectively.(2) All models contain time dummies for 1996 through 2001 and the FPE and the MWNLS estimator the time averages of the explanatory variables. (3) The numbers in parenthesis are robust standard errors. (4) The standard errors for the APEs are bootstrapped standard errors from 500 replications. (5) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table A.3: The Fractional Time-varying Model: Exogenous Spending

Model	Fractional model with time-varying heterogeneity					
Method	CMD		CFPE		CMWNLS	
Dependent: Test pass rate	Coef	APE	Coef	APE	Coef	APE
η_2	0.844		0.865		0.863	
	(0.091)		(0.059)		(0.06)	
η_3	0.885		0.904		0.904	
	(0.084)		(0.052)		(0.052)	
η_4	1.08		1.069		1.071	
	(0.106)		(0.076)		(0.077)	
η_5	1.225		1.232		1.244	
	(0.16)		(0.131)		(0.137)	
η_6	1.221		1.205		1.207	
	(0.162)		(0.121)		(0.122)	
η_7	1.113		1.127		1.132	
	(0.117)		(0.095)		(0.097)	
$\log(avgrexpp)$	0.949	0.214	0.997	0.225	1.013	0.228
	(0.248)	(0.064)	(0.242)	(0.064)	(0.244)	(0.067)
$lunch$	-0.282	-0.063	-0.332	-0.079	-0.355	-0.083
	(0.275)	(0.066)	(0.217)	(0.06)	(0.216)	(0.062)
$\log(enroll)$	0.339	0.08	0.357	0.082	0.356	0.082
	(0.169)	(0.027)	(0.146)	(0.025)	(0.147)	(0.026)
Minimum distance	35.000	(0.000)				

Notes: (1) CMD, CFPE, and CMWNSL stand for the classic minimum distance estimator, the constrained fractional probit estimator, and the constrained multivariate weighted nonlinear least squares estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t. (3) All models contain time dummies for 1996 through 2001 and the time averages of the explanatory variables. (4) The numbers in parenthesis are robust standard errors. (5) The standard errors for the APEs are bootstrapped standard errors from 500 replications. (6) Minimum distance statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table A.4: The Linear and the Fractional Model: Endogenous Spending

Model	Linear	Fractional			
Method	FEIV	Two-stage FPE		QLIMLE	
Dependent: Test pass rate	Coef	Coef	APE	Coef	APE
log(*avgrexpp*)	0.63	1.714	0.585	2.124	0.767
	(0.202)	(0.716)	(0.241)	(0.184)	(0.334)
lunch	-0.051	-0.288	-0.098	-0.319	-0.111
	(0.074)	(0.209)	(0.071)	(0.177)	(0.073)
log(*enroll*)	0.036	0.284	0.097	0.376	0.138
	(0.059)	(0.218)	(0.074)	(0.173)	(0.088)
$\hat{r}_2$	-0.49	-1.352		-1.785	
	(0.227)	(0.773)		(0.307)	
LM statistics		21.049	(0.002)	22.857	(0.001)
Variable addition		16.57	(0.011)		

Notes: (1) FEIV, FPE, and QLIMLE stand for the fixed effects instrumental variable estimator, the fractional probit estimator, and the quasi limited information maximum likelihood estimator, respectively. (2) All models contain time dummies for 1996 through 2001, the time averages of lunch and log(enroll), log of spending per student in 1994, and interaction terms between this variable and time dummies for 1996 through 2001. (3) The instrument variables are the log of foundation grant and interaction terms between this variable and time dummies for 1996 through 2001. (4) The numbers in parenthesis are bootstrapped standard errors from 500 replications. (5) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table A.5: The Fractional Time-varying Model: Endogenousspending, Two-step Estimator

Method	Two-stage CMD				Two-stage FPE			
Dependent: Test pass rates	Coef.	Std Err	APE	Std Err	Coef.	Std Err	APE	Std Err
η_2	0.891	(0.086)			0.959	(0.057)		
η_3	0.926	(0.086)			0.968	(0.057)		
η_4	1.079	(0.124)			1.077	(0.073)		
η_5	1.205	(0.219)			1.254	(0.187)		
η_6	1.129	(0.167)			1.190	(0.183)		
η_7	1.151	(0.209)			1.140	(0.128)		
$\log(avgrexpp)$	2.064	(0.662)	0.507	(0.166)	2.502	(0.776)	0.592	(0.193)
$lunch$	-0.132	(0.303)	-0.032	(0.074)	-0.363	(0.361)	-0.086	(0.085)
$\log(enroll)$	0.431	(0.2)	0.105	(0.048)	0.592	(0.237)	0.140	(0.056)
φ_1	-1.984	(0.893)			-2.293	(0.985)		
φ_2	-1.986	(0.881)			-2.235	(1.02)		
φ_3	-1.434	(0.956)			-2.521	(1.059)		
φ_4	-2.403	(0.965)			-2.909	(1.012)		
φ_5	-0.533	(1.048)			-1.044	(1.12)		
φ_6	-1.396	(0.854)			-1.805	(1.002)		
φ_7	-0.813	(0.919)			-1.343	(0.971)		
Minimum distance	49.182	(0.000)						
Wald statistics	17.933	(0.012)			18.553	(0.010)		

Notes: (1) CMD and CFPE stand for the classic minimum distance estimator and the constrained fractional probit estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t and φ_t is the coefficient of the residual or error term from the reduced form. (3) All models contain time dummies for 1996 through 2001, the time averages of lunch and log(enroll), log of spending per student in 1994, and interaction terms between this variable and time dummies for 1996 through 2001. (4) The numbers in parenthesis are bootstrapped standard errors from 500 replications. (5) Minimum distance statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values. (6) Wald test is for test endogeneity of the school spending ($H_0 : \varphi_t = 0$) and the number in parenthesis are the corresponding p-value.

Table A.6: The Fractional Time-varying Model: Endogenous Spending, Limited Information MLE

Method	LIML-CMD				CQLIMLE			
Dependent: Test pass rates	Coef.	Std Err	APE	Std Err	Coef.	Std Err	APE	Std Err
η_2	0.932	(0.061)			0.945	(0.067)		
η_3	0.940	(0.067)			0.953	(0.067)		
η_4	1.062	(0.084)			1.069	(0.081)		
η_5	1.240	(0.198)			1.259	(0.152)		
η_6	1.191	(0.19)			1.188	(0.127)		
η_7	1.157	(0.167)			1.140	(0.108)		
$\log(avgrexpp)$	2.484	(0.639)	0.566	(0.149)	2.490	(0.402)	0.570	(0.09)
lunch	-0.453	(0.429)	-0.102	(0.096)	-0.642	(0.349)	-0.148	(0.081)
$\log(enroll)$	0.568	(0.233)	0.129	(0.052)	0.546	(0.239)	0.125	(0.052)
φ_1	-2.474	(0.99)			-2.409	(0.749)		
φ_2	-2.336	(1.015)			-2.436	(0.807)		
φ_3	-2.621	(0.953)			-2.517	(0.736)		
φ_4	-2.999	(0.957)			-2.899	(0.7)		
φ_5	-0.956	(1.039)			-0.915	(0.776)		
φ_6	-1.720	(0.913)			-1.647	(0.696)		
φ_7	-1.260	(0.85)			-1.238	(0.644)		
Minimum distance	29.767	(0.000)						
Wald statistics	21.503	(0.003)			18.062	(0.012)		

Notes: (1) LIME-CMD and CQLIMLE stand for the limited information maximum likelihood estimator-classic minimum distance estimator and the constrained quasi limited information maximum likelihood estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t and φ_t is the coefficient of the residual or error term from the reduced form. (3) All models contain time dummies for 1996 through 2001, the time averages of lunch and log(enroll), log of spending per student in 1994, and interaction terms between this variable and time dummies for 1996 through 2001. (4) The numbers in parenthesis are bootstrapped standard errors from 500 replications. (5) Minimum distance statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values. (6) Wald test is for test endogeneity of the school spending ($H_0 : \varphi_t = 0$) and the number in parenthesis are the corresponding p-value.

Table A.7: The Scale Factors at Different Spending Levels (in 1995 and 2001)

	Papke and Wooldridge (2008)				Time-varying model							
	Exogenous Spending											
	QMLE				CMD				Constrained MWNLS			
Percentile	APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}	
5th	0.342	(0.081)	0.293	(0.073)	0.268	(0.075)	0.212	(0.069)	0.263	(0.072)	0.209	(0.065)
25th	0.338	(0.08)	0.284	(0.069)	0.265	(0.074)	0.206	(0.066)	0.261	(0.071)	0.203	(0.063)
50th	0.333	(0.079)	0.277	(0.065)	0.262	(0.073)	0.202	(0.064)	0.257	(0.069)	0.199	(0.06)
75th	0.322	(0.073)	0.265	(0.06)	0.254	(0.069)	0.195	(0.06)	0.25	(0.066)	0.192	(0.057)
95th	0.293	(0.058)	0.229	(0.043)	0.235	(0.058)	0.174	(0.048)	0.232	(0.056)	0.172	(0.046)
	Endogenous Spending (Two-step procedure)											
	QMLE				CMD				Constrained FPE			
Percentile	APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}	
5th	0.653	(0.248)	0.602	(0.258)	0.559	(0.171)	0.455	(0.172)	0.670	(0.186)	0.569	(0.208)
25th	0.653	(0.255)	0.571	(0.24)	0.555	(0.173)	0.434	(0.16)	0.668	(0.192)	0.540	(0.193)
50th	0.64	(0.249)	0.542	(0.221)	0.544	(0.168)	0.417	(0.149)	0.654	(0.188)	0.515	(0.177)
75th	0.589	(0.211)	0.491	(0.183)	0.509	(0.148)	0.386	(0.13)	0.605	(0.158)	0.470	(0.147)
95th	0.433	(0.106)	0.341	(0.085)	0.406	(0.091)	0.296	(0.077)	0.454	(0.078)	0.339	(0.068)
	Endogenous Spending (LIML)											
	QMLE				CMD				Constrained FPE			
per.	APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}		APE_{1995}		APE_{2001}	
5th	0.827	(0.316)	0.797	(0.348)	0.668	(0.164)	0.526	(0.166)	0.672	(0.098)	0.544	(0.1)
25th	0.837	(0.334)	0.752	(0.325)	0.665	(0.166)	0.501	(0.154)	0.669	(0.101)	0.518	(0.092)
50th	0.819	(0.327)	0.706	(0.297)	0.651	(0.162)	0.479	(0.143)	0.655	(0.098)	0.495	(0.085)
75th	0.732	(0.268)	0.62	(0.241)	0.604	(0.141)	0.440	(0.122)	0.609	(0.083)	0.456	(0.072)
95th	0.471	(0.128)	0.376	(0.111)	0.456	(0.076)	0.327	(0.065)	0.468	(0.042)	0.342	(0.036)

Notes: (1) These scale factors are obtained from table (A.2)- table (A.6). (2) The numbers in parenthesis are bootstrapped standard errors from 500 replications.

Table A.8: Mean Squared Errors and Biases of the APEs (Increasing η, $\rho_{xc} = 0.1 - 0.5$)

ρ_{xc}	0.1		0.2		0.3		0.4		0.5	
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
					T=5					
Linear	0.0115	-0.0024	0.0140	-0.0083	0.0176	-0.0136	0.0216	-0.0184	0.0255	-0.0228
PW_FPE	0.0112	-0.0025	0.0138	-0.0084	0.0175	-0.0137	0.0216	-0.0186	0.0255	-0.0230
PW_MWNLS	0.0112	-0.0026	0.0138	-0.0084	0.0176	-0.0137	0.0216	-0.0186	0.0256	-0.0231
TV_CMD	0.0102	-0.0025	0.0131	-0.0083	0.0168	-0.0137	0.0208	-0.0186	0.0247	-0.0232
TV_CFPE	0.0104	-0.0026	0.0132	-0.0084	0.0168	-0.0137	0.0208	-0.0187	0.0248	-0.0233
TV_CMWNLS	0.0061	-0.0024	0.0103	-0.0084	0.0152	-0.0138	0.0199	-0.0188	0.0243	-0.0234
					T=7					
Linear	0.0179	-0.0098	0.0219	-0.0161	0.0263	-0.0217	0.0307	-0.0268	0.0348	-0.0314
PW_FPE	0.0178	-0.0099	0.0219	-0.0161	0.0264	-0.0218	0.0307	-0.0269	0.0348	-0.0315
PW_MWNLS	0.0178	-0.0099	0.0219	-0.0162	0.0264	-0.0218	0.0307	-0.0269	0.0348	-0.0316
TV_CMD	0.0146	-0.0098	0.0193	-0.0161	0.0240	-0.0219	0.0286	-0.0272	0.0332	-0.0321
TV_CFPE	0.0147	-0.0099	0.0193	-0.0161	0.0240	-0.0219	0.0286	-0.0272	0.0331	-0.0321
TV_CMWNLS	0.0111	-0.0097	0.0171	-0.0162	0.0228	-0.0220	0.0280	-0.0274	0.0329	-0.0323
					T=10					
Linear	0.0202	-0.0035	0.0218	-0.0090	0.0242	-0.0138	0.0270	-0.0182	0.0297	-0.0220
PW_FPE	0.0202	-0.0037	0.0219	-0.0091	0.0243	-0.0139	0.0270	-0.0182	0.0297	-0.0221
PW_MWNLS	0.0202	-0.0037	0.0219	-0.0091	0.0243	-0.0140	0.0271	-0.0184	0.0298	-0.0222
TV_CMD	0.0118	-0.0035	0.0138	-0.0090	0.0170	-0.0141	0.0208	-0.0188	0.0246	-0.0229
TV_CFPE	0.0119	-0.0036	0.0138	-0.0091	0.0169	-0.0141	0.0207	-0.0188	0.0245	-0.0229
TV_CMWNLS	0.0068	-0.0034	0.0109	-0.0091	0.0157	-0.0144	0.0203	-0.0191	0.0245	-0.0232

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator. (2) ρ_{xc} is the population correlation between the explanatory variable (x) and individual heterogeneity (c).

Table A.9: Mean Squared Errors and Biases of the APEs (Increasing η, $\rho_{xc} = 0.6 - 0.9$)

ρ_{xc}	0.6		0.7		0.8		0.9	
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
				T=5				
Linear	0.0291	-0.0268	0.0325	-0.0304	0.0357	-0.0337	0.0385	-0.0367
PW_FPE	0.0293	-0.0271	0.0327	-0.0308	0.0360	-0.0341	0.0388	-0.0372
PW_MWNLS	0.0293	-0.0271	0.0328	-0.0308	0.0360	-0.0342	0.0388	-0.0372
TV_CMD	0.0286	-0.0274	0.0322	-0.0312	0.0357	-0.0349	0.0398	-0.0389
TV_CFPE	0.0286	-0.0275	0.0323	-0.0314	0.0359	-0.0351	0.0398	-0.0390
TV_CMWNLS	0.0284	-0.0276	0.0323	-0.0315	0.0360	-0.0352	0.0402	-0.0392
				T=7				
Linear	0.0385	-0.0355	0.0420	-0.0393	0.0450	-0.0425	0.0477	-0.0454
PW_FPE	0.0386	-0.0357	0.0421	-0.0395	0.0452	-0.0427	0.0478	-0.0455
PW_MWNLS	0.0387	-0.0357	0.0421	-0.0395	0.0452	-0.0428	0.0478	-0.0455
TV_CMD	0.0374	-0.0365	0.0415	-0.0407	0.0454	-0.0446	0.0508	-0.0498
TV_CFPE	0.0374	-0.0365	0.0414	-0.0406	0.0453	-0.0445	0.0499	-0.0487
TV_CMWNLS	0.0374	-0.0367	0.0416	-0.0408	0.0456	-0.0446	0.0503	-0.0486
				T=10				
Linear	0.0323	-0.0255	0.0348	-0.0285	0.0371	-0.0313	0.0392	-0.0337
PW_FPE	0.0323	-0.0254	0.0347	-0.0284	0.0368	-0.0310	0.0387	-0.0333
PW_MWNLS	0.0323	-0.0255	0.0347	-0.0284	0.0368	-0.0310	0.0387	-0.0333
TV_CMD	0.0281	-0.0267	0.0316	-0.0302	0.0356	-0.0341	0.0419	-0.0402
TV_CFPE	0.0282	-0.0267	0.0318	-0.0301	0.0355	-0.0335	0.0403	-0.0375
TV_CMWNLS	0.0284	-0.0269	0.0321	-0.0301	0.0360	-0.0333	0.0407	-0.0368

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator. (2) ρ_{xc} is the population correlation between the explanatory variable (x) and individual heterogeneity (c).

Table A.10: Mean Squared Errors and Biases of the APEs (Decreasing η, $\rho_{xc} = 0.1 - 0.5$)

ρ_{xc}	0.1		0.2		0.3		0.4		0.5	
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
T=5										
Linear	0.0113	-0.0041	0.0145	-0.0099	0.0185	-0.0152	0.0227	-0.0201	0.0268	-0.0246
PW_FPE	0.0114	-0.0042	0.0146	-0.0100	0.0187	-0.0154	0.0229	-0.0203	0.0270	-0.0248
PW_MWNLS	0.0114	-0.0042	0.0146	-0.0100	0.0187	-0.0154	0.0229	-0.0203	0.0270	-0.0248
TV_CMD	0.0085	-0.0042	0.0125	-0.0101	0.0176	-0.0160	0.0231	-0.0218	0.0289	-0.0275
TV_CFPE	0.0086	-0.0042	0.0125	-0.0101	0.0172	-0.0156	0.0220	-0.0208	0.0268	-0.0257
TV_CMWNLS	0.0065	-0.0042	0.0115	-0.0102	0.0166	-0.0157	0.0217	-0.0208	0.0265	-0.0255
T=7										
Linear	0.0157	-0.0080	0.0196	-0.0141	0.0240	-0.0198	0.0285	-0.0251	0.0328	-0.0299
PW_FPE	0.0160	-0.0081	0.0198	-0.0142	0.0242	-0.0199	0.0286	-0.0251	0.0329	-0.0299
PW_MWNLS	0.0160	-0.0081	0.0198	-0.0143	0.0242	-0.0199	0.0286	-0.0251	0.0329	-0.0299
TV_CMD	0.0101	-0.0078	0.0163	-0.0149	0.0234	-0.0219	0.0304	-0.0284	0.0368	-0.0345
TV_CFPE	0.0102	-0.0079	0.0158	-0.0144	0.0219	-0.0205	0.0278	-0.0260	0.0333	-0.0311
TV_CMWNLS	0.0090	-0.0080	0.0154	-0.0144	0.0216	-0.0204	0.0275	-0.0257	0.0329	-0.0307
T=10										
Linear	0.0174	-0.0047	0.0193	-0.0094	0.0218	-0.0138	0.0245	-0.0178	0.0273	-0.0214
PW_FPE	0.0178	-0.0048	0.0196	-0.0095	0.0219	-0.0137	0.0245	-0.0176	0.0271	-0.0212
PW_MWNLS	0.0178	-0.0048	0.0196	-0.0095	0.0219	-0.0138	0.0245	-0.0177	0.0272	-0.0212
TV_CMD	0.0073	-0.0052	0.0133	-0.0112	0.0203	-0.0173	0.0270	-0.0231	0.0327	-0.0282
TV_CFPE	0.0071	-0.0048	0.0122	-0.0099	0.0179	-0.0145	0.0232	-0.0188	0.0278	-0.0226
TV_CMWNLS	0.0063	-0.0048	0.0119	-0.0098	0.0176	-0.0143	0.0228	-0.0183	0.0273	-0.0219

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator. (2) ρ_{xc} is the population correlation between the explanatory variable (x) and individual heterogeneity (c).

Table A.11: Mean Squared Errors and Biases of the APEs (Decreasing η, $\rho_{xc} = 0.6 - 0.9$)

ρ_{xc}	0.6		0.7		0.8		0.9	
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
T=5								
Linear	0.0306	-0.0287	0.0343	-0.0325	0.0376	-0.0360	0.0407	-0.0392
PW_FPE	0.0308	-0.0290	0.0345	-0.0328	0.0378	-0.0363	0.0410	-0.0396
PW_MWNLS	0.0309	-0.0290	0.0345	-0.0328	0.0379	-0.0364	0.0410	-0.0396
TV_CMD	0.0345	-0.0328	0.0397	-0.0376	0.0446	-0.0421	0.0492	-0.0462
TV_CFPE	0.0315	-0.0302	0.0361	-0.0345	0.0407	-0.0384	0.0452	-0.0421
TV_CMWNLS	0.0311	-0.0298	0.0356	-0.0339	0.0401	-0.0378	0.0448	-0.0416
T=7								
Linear	0.0368	-0.0342	0.0405	-0.0382	0.0439	-0.0418	0.0469	-0.0450
PW_FPE	0.0369	-0.0343	0.0406	-0.0383	0.0439	-0.0419	0.0468	-0.0450
PW_MWNLS	0.0369	-0.0343	0.0406	-0.0383	0.0439	-0.0419	0.0468	-0.0450
TV_CMD	0.0426	-0.0399	0.0478	-0.0448	0.0524	-0.0492	0.0564	-0.0529
TV_CFPE	0.0384	-0.0358	0.0430	-0.0400	0.0472	-0.0438	0.0511	-0.0472
TV_CMWNLS	0.0378	-0.0352	0.0424	-0.0393	0.0467	-0.0432	0.0508	-0.0467
T=10								
Linear	0.0299	-0.0247	0.0325	-0.0276	0.0349	-0.0303	0.0370	-0.0327
PW_FPE	0.0297	-0.0244	0.0321	-0.0272	0.0343	-0.0298	0.0363	-0.0321
PW_MWNLS	0.0297	-0.0244	0.0321	-0.0273	0.0343	-0.0298	0.0363	-0.0321
TV_CMD	0.0376	-0.0326	0.0417	-0.0364	0.0453	-0.0396	0.0483	-0.0419
TV_CFPE	0.0319	-0.0260	0.0356	-0.0290	0.0390	-0.0317	0.0422	-0.0341
TV_CMWNLS	0.0314	-0.0252	0.0351	-0.0282	0.0385	-0.0310	0.0419	-0.0337

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator. (2) ρ_{xc} is the population correlation between the explanatory variable (x) and individual heterogeneity (c).

Table A.12: Mean Squared Errors and Biases of the APEs ($\rho_{xc} = 0.2$)

	Increasing η						Decreasing η					
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
N	100		300		500		100		300		500	
							T=5					
Linear	0.0135	0.0005	0.0142	-0.0082	0.0140	-0.0083	0.0139	-0.0042	0.0147	-0.0097	0.0145	-0.0099
PW_FPE	0.0126	0.0000	0.0140	-0.0084	0.0138	-0.0084	0.0126	-0.0047	0.0147	-0.0100	0.0146	-0.0100
PW_MWNLS	0.0124	-0.0001	0.0139	-0.0084	0.0138	-0.0084	0.0125	-0.0048	0.0147	-0.0101	0.0146	-0.0100
TV_CMD	0.0207	0.0006	0.0152	-0.0083	0.0131	-0.0083	0.0156	-0.0041	0.0134	-0.0099	0.0125	-0.0101
TV_CFPE	0.0207	0.0001	0.0153	-0.0085	0.0132	-0.0084	0.0156	-0.0046	0.0135	-0.0100	0.0125	-0.0101
TV_CMWNLS	0.0139	0.0002	0.0114	-0.0084	0.0103	-0.0084	0.0123	-0.0047	0.0120	-0.0101	0.0115	-0.0102
							T=7					
Linear	0.0190	-0.0099	0.0202	-0.0133	0.0219	-0.0161	0.0193	-0.0110	0.0190	-0.0128	0.0196	-0.0141
PW_FPE	0.0188	-0.0103	0.0202	-0.0135	0.0219	-0.0161	0.0186	-0.0116	0.0190	-0.0130	0.0198	-0.0142
PW_MWNLS	0.0187	-0.0104	0.0202	-0.0136	0.0219	-0.0162	0.0185	-0.0117	0.0190	-0.0130	0.0198	-0.0143
TV_CMD	0.0240	-0.0099	0.0189	-0.0134	0.0193	-0.0161	0.0173	-0.0119	0.0158	-0.0137	0.0163	-0.0149
TV_CFPE	0.0241	-0.0101	0.0189	-0.0135	0.0193	-0.0161	0.0171	-0.0116	0.0153	-0.0131	0.0158	-0.0144
TV_CMWNLS	0.0163	-0.0101	0.0153	-0.0135	0.0171	-0.0162	0.0149	-0.0117	0.0145	-0.0132	0.0154	-0.0144
							T=10					
Linear	0.0314	-0.0099	0.0232	-0.0133	0.0218	-0.0090	0.0257	-0.0175	0.0205	-0.0118	0.0193	-0.0094
PW_FPE	0.0310	-0.0103	0.0233	-0.0135	0.0219	-0.0091	0.0248	-0.0174	0.0204	-0.0118	0.0196	-0.0095
PW_MWNLS	0.0310	-0.0104	0.0233	-0.0136	0.0219	-0.0091	0.0248	-0.0174	0.0204	-0.0119	0.0196	-0.0095
TV_CMD	0.0333	-0.0099	0.0180	-0.0134	0.0138	-0.0090	0.0226	-0.0185	0.0151	-0.0117	0.0133	-0.0112
TV_CFPE	0.0332	-0.0101	0.0180	-0.0135	0.0138	-0.0091	0.0219	-0.0174	0.0142	-0.0118	0.0122	-0.0099
TV_CMWNLS	0.0264	-0.0101	0.0141	-0.0135	0.0109	-0.0091	0.0209	-0.0172	0.0138	-0.0119	0.0119	-0.0098

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator.

Table A.13: Mean Squared Errors and Biases of the APEs ($\rho_{xc} = 0.3$)

	Increasing η						Decreasing η					
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
N	100		300		500		100		300		500	
T=5												
Linear	0.0142	-0.0041	0.0179	-0.0135	0.0176	-0.0136	0.0160	-0.0088	0.0187	-0.0151	0.0185	-0.0152
PW_FPE	0.0136	-0.0046	0.0177	-0.0138	0.0175	-0.0137	0.0150	-0.0093	0.0188	-0.0154	0.0187	-0.0154
PW_MWNLS	0.0134	-0.0047	0.0177	-0.0138	0.0176	-0.0137	0.0150	-0.0094	0.0188	-0.0154	0.0187	-0.0154
TV_CMD	0.0203	-0.0042	0.0182	-0.0137	0.0168	-0.0137	0.0175	-0.0091	0.0180	-0.0158	0.0176	-0.0160
TV_CFPE	0.0203	-0.0045	0.0183	-0.0139	0.0168	-0.0137	0.0173	-0.0093	0.0178	-0.0156	0.0172	-0.0156
TV_CMWNLS	0.0147	-0.0047	0.0160	-0.0139	0.0152	-0.0138	0.0150	-0.0095	0.0170	-0.0156	0.0166	-0.0157
T=7												
Linear	0.0222	-0.0150	0.0241	-0.0187	0.0263	-0.0217	0.0225	-0.0160	0.0229	-0.0182	0.0240	-0.0198
PW_FPE	0.0222	-0.0156	0.0242	-0.0190	0.0264	-0.0218	0.0222	-0.0167	0.0230	-0.0184	0.0242	-0.0199
PW_MWNLS	0.0222	-0.0157	0.0243	-0.0191	0.0264	-0.0218	0.0222	-0.0167	0.0230	-0.0184	0.0242	-0.0199
TV_CMD	0.0254	-0.0154	0.0226	-0.0190	0.0240	-0.0219	0.0221	-0.0185	0.0223	-0.0204	0.0234	-0.0219
TV_CFPE	0.0254	-0.0154	0.0226	-0.0190	0.0240	-0.0219	0.0209	-0.0171	0.0207	-0.0189	0.0219	-0.0205
TV_CMWNLS	0.0202	-0.0157	0.0206	-0.0192	0.0228	-0.0220	0.0197	-0.0169	0.0203	-0.0188	0.0216	-0.0204
T=10												
Linear	0.0350	-0.0282	0.0260	-0.0166	0.0242	-0.0138	0.0293	-0.0226	0.0233	-0.0157	0.0218	-0.0138
PW_FPE	0.0346	-0.0281	0.0261	-0.0166	0.0243	-0.0139	0.0284	-0.0223	0.0231	-0.0155	0.0219	-0.0137
PW_MWNLS	0.0348	-0.0283	0.0262	-0.0168	0.0243	-0.0140	0.0284	-0.0223	0.0231	-0.0155	0.0219	-0.0138
TV_CMD	0.0350	-0.0281	0.0209	-0.0168	0.0170	-0.0141	0.0296	-0.0254	0.0222	-0.0190	0.0203	-0.0173
TV_CFPE	0.0348	-0.0280	0.0208	-0.0168	0.0169	-0.0141	0.0275	-0.0228	0.0199	-0.0163	0.0179	-0.0145
TV_CMWNLS	0.0311	-0.0285	0.0189	-0.0172	0.0157	-0.0144	0.0270	-0.0224	0.0197	-0.0160	0.0176	-0.0143

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator.

Table A.14: Mean Squared Errors and Biases of the APEs ($\rho_{xc} = 0.4$)

	Increasing η						Decreasing η					
Model/Method	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
N	100		300		500		100		300		500	
T=5												
Linear	0.0160	-0.0081	0.0218	-0.0184	0.0216	-0.0184	0.0186	-0.0128	0.0230	-0.0201	0.0227	-0.0201
PW_FPE	0.0156	-0.0088	0.0218	-0.0187	0.0216	-0.0186	0.0180	-0.0135	0.0231	-0.0204	0.0229	-0.0203
PW_MWNLS	0.0155	-0.0089	0.0218	-0.0187	0.0216	-0.0186	0.0180	-0.0136	0.0231	-0.0204	0.0229	-0.0203
TV_CMD	0.0206	-0.0085	0.0218	-0.0188	0.0208	-0.0186	0.0203	-0.0139	0.0233	-0.0216	0.0231	-0.0218
TV_CFPE	0.0207	-0.0088	0.0219	-0.0189	0.0208	-0.0187	0.0197	-0.0137	0.0224	-0.0208	0.0220	-0.0208
TV_CMWNLS	0.0166	-0.0091	0.0205	-0.0190	0.0199	-0.0188	0.0182	-0.0138	0.0219	-0.0207	0.0217	-0.0208
T=7												
Linear	0.0256	-0.0197	0.0281	-0.0236	0.0307	-0.0268	0.0260	-0.0206	0.0270	-0.0231	0.0285	-0.0251
PW_FPE	0.0258	-0.0203	0.0283	-0.0239	0.0307	-0.0269	0.0259	-0.0213	0.0271	-0.0233	0.0286	-0.0251
PW_MWNLS	0.0259	-0.0204	0.0284	-0.0240	0.0307	-0.0269	0.0259	-0.0214	0.0271	-0.0233	0.0286	-0.0251
TV_CMD	0.0276	-0.0204	0.0267	-0.0242	0.0286	-0.0272	0.0276	-0.0246	0.0288	-0.0267	0.0304	-0.0284
TV_CFPE	0.0275	-0.0204	0.0266	-0.0242	0.0286	-0.0272	0.0253	-0.0221	0.0263	-0.0242	0.0278	-0.0260
TV_CMWNLS	0.0243	-0.0208	0.0256	-0.0244	0.0280	-0.0274	0.0245	-0.0217	0.0258	-0.0239	0.0275	-0.0257
T=10												
Linear	0.0381	-0.0321	0.0288	-0.0207	0.0270	-0.0182	0.0381	-0.0269	0.0288	-0.0197	0.0245	-0.0178
PW_FPE	0.0379	-0.0320	0.0290	-0.0209	0.0270	-0.0182	0.0379	-0.0265	0.0290	-0.0195	0.0245	-0.0176
PW_MWNLS	0.0381	-0.0324	0.0292	-0.0211	0.0271	-0.0184	0.0381	-0.0266	0.0292	-0.0195	0.0245	-0.0177
TV_CMD	0.0372	-0.0324	0.0243	-0.0214	0.0208	-0.0188	0.0372	-0.0316	0.0243	-0.0249	0.0270	-0.0231
TV_CFPE	0.0370	-0.0323	0.0242	-0.0214	0.0207	-0.0188	0.0370	-0.0274	0.0242	-0.0206	0.0232	-0.0188
TV_CMWNLS	0.0351	-0.0329	0.0233	-0.0219	0.0203	-0.0191	0.0351	-0.0268	0.0233	-0.0201	0.0228	-0.0183

Notes: (1) PW and TV stand for Papke and Wooldridge (2008) and the fractional model with time-varying individual heterogeneity. FPE, MWNSL, CMD, CFPE, CMWNLS stand for the fractional probit estimator, the multivariate weighted nonlinear squares estimator, the classical minimum distance estimator, the constrained FPE and the constrained MWNLS estimator.

Appendix B Tables for Chapter 2

Table B.1: Descriptive Statistics

Variable	Description	1986	1987	1988	1989	Min	Max
fwh	working hours/total hours	0.19	0.194	0.196	0.202	0	1
		(0.151)	(0.152)	(0.154)	(0.151)		
fert	1 if the age of the youngest baby is 1 at t+1	0.143	0.11	0.112	0.09	0	1
		(0.35)	(0.313)	(0.315)	(0.286)		
black	1 if the mother is black	0.773	0.773	0.773	0.773	0	1
		(0.419)	(0.419)	(0.419)	(0.419)		
age	the mother's age	31.302	32.302	33.302	34.302	19	57
		(5.698)	(5.698)	(5.698)	(5.698)		
*kid*26	1 if the mother has kids aged 2 - 6	0.37	0.377	0.379	0.384	0	1
		(0.483)	(0.485)	(0.485)	(0.487)		
hinc	log(husband's income)	11.812	12.391	13.207	14.023	0	99.99
		(8.521)	(8.652)	(9.808)	(10.507)		
dsex	1 if the first two children have the same sex	0.234	0.246	0.262	0.283	0	1
		(0.424)	(0.431)	(0.44)	(0.451)		

Note: the numbers are mean values and the numbers in parenthesis are standard deviations.

Table B.2: The Linear and the Fractional Model: Exogenous Fertility

Model	Linear	Fractional			
Method	Fixed effects	Pooled FPE		MWNLS	
Dependent: fwh	Coef	Coef	APE	Coef	APE
black	-0.020	-0.073		-0.073	
	(0.009)	(0.03)		(0.03)	
age	0.000	0.000		0.000	
	(0.001)	(0.002)		(0.002)	
$fert$	-0.021	-0.087	-0.023	-0.087	-0.023
	(0.004)	(0.017)	(0.004)	(0.017)	(0.004)
$kid26$	0.001	0.005	0.001	0.005	0.001
	(0.004)	(0.014)	(0.004)	(0.014)	(0.004)
hinc	-0.001	-0.004	-0.001	-0.004	-0.001
	(0)	(0.001)	(0.000)	(0.001)	(0.000)
LM (score) test		2.4957	(0.4761)	2.4943	(0.4763)
Variable addition test		8.1776	(0.0425)	8.2597	(0.0409)

Notes: (1) FPE, and MWNSL stand for the fractional probit estimator, and multivariate weighted nonlinear least squares estimator, respectively. (2) all models contains year dummies for 1986 through 1989 and the fractional model has the time averages of $fert$, $kid26$, and *hinc*.(3) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (4) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table B.3: The Fractional Time-varying Model: Exogenous Fertility

Method	pooled CFPE		CMD		CMWNLS	
Dependent: fwh	Coef	APE	Coef	APE	Coef	APE
η_2	1.425		1.449		1.485	
	(0.591)		(0.607)		(0.349)	
η_3	1.075		1.093		1.120	
	(0.357)		(0.369)		(0.232)	
η_4	0.867		0.849		0.863	
	(0.255)		(0.249)		(0.163)	
$black$	-0.108		-0.107		-0.110	
	(0.027)		(0.027)		(0.041)	
age	0.000		0.000		0.000	
	(0.002)		(0.002)		(0.003)	
$fert$	-0.125	-0.022	-0.127	-0.023	-0.127	-0.023
	(0.048)	(0.005)	(0.047)	(0.005)	(0.027)	(0.004)
$kid26$	0.014	0.003	0.015	0.003	0.017	0.003
	(0.035)	(0.005)	(0.035)	(0.004)	(0.021)	(0.004)
$hinc$	-0.006	-0.001	-0.006	-0.001	-0.006	-0.001
	(0.003)	(0.000)	(0.003)	(0.000)	(0.002)	(0.000)
GMM test			2.6018	(0.4572)		

Notes: (1) CFPE, CMD, and CMWNSL stand for the constrained fractional probit estimator, the classical minimum distance estimator, and the constrained multivariate weighted nonlinear least squares estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t. (3) all models contains year dummies for 1986 through 1989 and the time averages of $fert$, $kid26$, and $hinc$. (4) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (5) GMM statistic is used to test $H_0 : \eta_t = 1$and the numbers in parenthesis are the corresponding p-values.

Table B.4: The Linear and Fractional Model: Endogenous Fertility

Model	Linear	Fractional	
Method	FEIV	Pooled FPE	
Dependent: fwh	Coefficient	Coefficient	APE
$black$	-0.020	-0.072	
	(0.009)	(0.017)	
age	0.000	0.002	
	(0.001)	(0.002)	
$fert$	-0.141	-0.581	-0.129
	(0.037)	(0.099)	(0.021)
$kid26$	-0.033	-0.142	-0.038
	(0.010)	(0.039)	(0.010)
$hinc$	-0.001	-0.003	-0.001
	(0.000)	(0.002)	(0.000)
ρ		0.214	
		(0.058)	
LM (score) test		1.4550	(0.6927)
Variable addition test		1.5245	(0.6766)

Notes: (1) FEIV and FPE stand for the fixed effects instrumental variable estimator and the fractional probit estimator, respectively. (2) all models contains year dummies for 1986 through 1989 and the fractional model has the time averages of $dsex$, $kid26$, and $hinc$.(3) ρ denotes the correlation between the error in the structural equation for fractional worked hours and the errors in the reduce form for fertility. (4) The numbers in parenthesis are robust standard errors and are obtained from 500 bootstrapping repetitions. (4) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table B.5: The Fractional Time-varying Model: Endogenous Fertility

Method	Pooled FBPE		CMD	
Dependent: fwh	Coef	APE	Coef	APE
η_2	1.043		0.968	
	(0.516)		(0.494)	
η_3	0.640		0.611	
	(0.327)		(0.317)	
η_4	0.392		0.287	
	(0.274)		(0.257)	
$black$	-0.099		-0.095	
	(0.028)		(0.028)	
age	0.003		0.002	
	(0.003)		(0.003)	
$fert$	-0.729	-0.122	-0.743	-0.126
	(0.181)	(0.024)	(0.177)	(0.025)
$kid26$	-0.159	-0.032	-0.164	-0.034
	(0.059)	(0.011)	(0.058)	(0.013)
$hinc$	-0.007	-0.001	-0.007	-0.002
	(0.003)	(0.000)	(0.003)	(0.000)
ρ_1	0.164			
	(0.081)			
Endogeneity test	10.7640	(0.0293)		
GMM test			4.2470	(0.2360)

Notes: (1) FBPE and CMD stand for the fractional bivariate probit estimator and the classical minimum distance estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t. (3) all models contains year dummies for 1986 through 1989 and the time averages of $dsex$, $kid26$, and $hinc$. (4) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (5) Endogeneity test is the Wald test for $H_0 : \rho_t = 0$. (6) GMM statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table B.6: Mean Squared Errors and Biases of the ATEs

Model	Linear		Fractional			
Method	FEIV		Two-step FPE		Single-step FBPE	
	MSE	Bias	MSE	Bias	MSE	Bias
N			$T = 5$			
300	0.07469	0.01379	0.06322	0.00226	0.05556	-0.00717
500	0.05517	0.01253	0.04718	0.00263	0.04207	-0.00581
700	0.04610	0.01323	0.03885	0.00366	0.03441	-0.00431
1000	0.03971	0.01129	0.03328	0.00188	0.02969	-0.00604
			$T = 7$			
300	0.05514	0.00584	0.04852	-0.00321	0.04486	-0.0106
500	0.04218	0.00897	0.03631	0.00017	0.03318	-0.00753
700	0.03757	0.00536	0.03319	-0.00268	0.03087	-0.00977
1000	0.02946	0.00672	0.02544	-0.00106	0.0238	-0.00774

Notes: (1) FEIV, FBE, and FBPE stand for the fixed effects instrumental variable estimator, the fractional probit estimator and the fractional bivariate probit estimator, respectively. (2) Two-step FPE plugs the fitted values for fertility obtained at the first stage as an additional variable and the single-step FBPE uses the modified bivariate probit model. (3) N and T denote the number of cross-sectional and time series observations.

Appendix C Tables for Chapter 3

Table C.1: The Fractional Hurdle Model: Exogenous Fertility

Method	Pooled QMLE			MWNLS		
Dependent: $1(fwk > 0)$	Coef			Coef		
$fert$	-0.160			-0.162		
	(0.044)			(0.044)		
$kid26$	0.002			0.001		
	(0.042)			(0.042)		
$hinc$	-0.004			-0.004		
	(0.004)			(0.004)		
Dependent: fwk	Coef	APE	APE $(fwk > 0)$	Coef	APE	APE $(fwk > 0)$
$fert$	-0.041	-0.022	-0.014	-0.041	-0.022	-0.013
	(0.018)	(0.004)	(0.005)	(0.018)	(0.004)	(0.005)
$kid26$	0.004	0.001	0.002	0.004	0.001	0.002
	(0.014)	(0.004)	(0.005)	(0.014)	(0.004)	(0.005)
$hinc$	-0.003	-0.001	-0.001	-0.003	-0.001	-0.001
	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
LM (score) test	1.7753	(0.6203)		1.8433	(0.6056)	
Variable addition	6.1924	(0.1026)		6.2496	(0.1001)	

Notes: (1) QMLE, and MWNSL stand for the quasi-maximum likelihood estimator, and multivariate weighted nonlinear least squares estimator, respectively. (2) all models contains year dummies for 1986 through 1989, black, age, and the time averages of $fert$, $kid26$, and $hinc$.(3) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (4) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table C.2: The Fractional Hurdle Time-varying Model: Exogenous Fertility

Method	Constrained QMLE		CMD		Constrained MWNSL	
Dependent: $1(fwk > 0)$	Coef	APE ($fwk > 0$)	Coef	APE ($fwk > 0$)	Coef	APE ($fwk > 0$)
$fert$	-0.162	-0.012		-0.011	-0.162	-0.013
	(0.067)	(0.006)		(0.007)	(0.043)	(0.006)
$kid26$	-0.002	0.005		0.004	0.000	0.005
	(0.059)	(0.006)		(0.006)	(0.041)	(0.005)
$hinc$	-0.004	-0.001		-0.001	-0.004	-0.001
	(0.005)	(0.000)		(0.000)	(0.003)	(0.000)
Dependent: fwk	Coef	APE	Coef	APE	Coef	APE
η_2	1.495		1.776		1.877	
	(0.992)		(1.335)		(0.771)	
η_3	0.909		1.041		1.036	
	(0.456)		(0.58)		(0.317)	
η_4	0.633		0.676		0.659	
	(0.298)		(0.341)		(0.204)	
$fert$	-0.050	-0.020	-0.056	-0.021	-0.054	-0.021
	(0.036)	(0.005)	(0.038)	(0.005)	(0.026)	(0.005)
$kid26$	0.025	0.004	0.018	0.003	0.028	0.005
	(0.028)	(0.004)	(0.029)	(0.004)	(0.021)	(0.004)
$hinc$	-0.005	-0.001	-0.005	-0.001	-0.005	-0.001
	(0.002)	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)
GMM test			2.2727	(0.5178)		

Notes: (1) QMLE, CMD, and MWNSL stand for the quasi-maximum likelihood estimator, the classical minimum distance estimator, and the multivariate weighted nonlinear least squares estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t. (3) all models contains year dummies for 1986 through 1989, black, age, and the time averages of $fert$, $kid26$, and $hinc$. (4) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (5) GMM statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table C.3: The Fractional Hurdle Model: Endogenous Fertility

Method	Quasi maximum likelihood estimator		
Dependent: $1(fwk > 0)$	Coef		
$fert$	-0.924		
	(0.273)		
$kid26$	-0.206		
	(0.092)		
$hinc$	-0.003		
	(0.005)		
ρ_1	0.301		
	(0.154)		
Dependent: fwk	Coef	APE	APE ($fwk > 0$)
$fert$	-0.334	-0.129	-0.091
	(0.129)	(0.027)	(0.026)
$kid26$	-0.076	-0.033	-0.022
	(0.039)	(0.009)	(0.009)
$hinc$	-0.003	-0.001	-0.0008
	(0.002)	(0.000)	(0.0006)
ρ_2	0.128		
	(0.078)		
LM (score) Test	0.4648	(0.9265)	
Variable addition	0.1742	(0.9816)	

Notes: (1) The model contains year dummies for 1986 through 1989, black, age, and the time averages of $dsex$, $kid26$, and $hinc$.(2) ρ_1 and ρ_2 denote the correlations between the error in equation for 1(fwk>0) and the errors in the reduce form for fertility, and the errors in equation for fwk and the errors in the reduce form for fertility. (3) The numbers in parenthesis are robust standard errors and are obtained from 500 bootstrapping repetitions. (4) LM statistics and variable addition test are used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

Table C.4: The Fractional Hurdle Time-varying Model: Endogenous Fertility

Method	Constrained QMLE			CMD		
Dependent: fwk	Coef	APE	APE ($fwk > 0$)	Coef	APE	APE ($fwk > 0$)
	(0.88)			(0.731)		
η_3	0.488			0.447		
	(0.454)			(0.429)		
η_4	0.037			0.054		
	(0.306)			(0.286)		
$fert$	-0.397	-0.124	-0.084	-0.308	-0.117	-0.071
	(0.188)	(0.022)	(0.028)	(0.18)	(0.025)	(0.026)
$kid26$	-0.070	-0.027	-0.018	-0.053	-0.025	-0.013
	(0.052)	(0.008)	(0.010)	(0.051)	(0.009)	(0.011)
$hinc$	-0.007	-0.002	-0.0013	-0.006	-0.003	-0.0015
	(0.003)	(0.002)	(0.0006)	(0.003)	(0.002)	(0.0005)
ρ_{21}	0.087					
	(0.085)					
Endogeneity test	2.3352	(0.6744)				
GMM test	4.0225	(0.2590)				

Notes: (1) QMLE and CMD stand for the quasi-maximum likelihood estimator and the classical minimum distance estimator, respectively. (2) η_t is the coefficient of individual heterogeneity for time t. (3) The model contains year dummies for 1986 through 1989, *black*, *age*, and the time averages of *dsex*, *kid26*, and *hinc*. (4) The numbers in parenthesis are robust standard errors and the standard errors of the APEs are obtained from 500 bootstrapping repetitions. (5) Endogeneity test is the Wald test for $H_0 : \rho_{2t} = 0$. (6) GMM statistic is used to test $H_0 : \eta_t = 1$ and the numbers in parenthesis are the corresponding p-values.

BIBLIOGRAPHY

Ahn S.C., Y.H. Lee and P. Schmidt (2001), "GMM estimation of linear panel data models with time-varying individual effects," *Journal of Econometrics* 101.

Ahn S.C, Y.H. Lee and P. Schmidt (2007), "Panel Data Models with Multiple Time-Varying Individual Effects," *Journal of Productivity Analysis* 27.

Angrist, D. and Evans, W. N. (1998), "Children and Their Parents Labor Supply: Evidence From Exogenous Variation in Family Size," *American Economic Review* 88.

Arellano, M. (1987), "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Econometrics and Statistics* 49.

Carrasco, R. (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.

Chamberlain, G. (1980), "Analysis of Variance with Qualitative Data," *Review of Economic Studies* 47.

Freedman, D. and Sekhon, J (2008), "Endogeneity in Probit Response Models," Unpublished. U. C. Berkeley CA 94720

Jacobsen, J.P., Pearce, J. and Rosenbloom, J.L (1999), "The Effects of Childbearing on Married Women's Labor Supply and Earnings Using Twin Births as a Natural Experiment," *The Journal of Human Resources* 34

Bay, J. (2005), "Panel Data Models with Interactive Fixed Effects," unpublished manuscript, New York University.

Kiefer, N.M. (1980), "A Time Series-Cross Section Model with Fixed Effects with an Intertemporal Factor Structure," unpublished manuscript, Cornel University.

Lee, Y.H. (1991), "Panel Data Models with Multiplicative Individual and Time Effects: Application to Compensation and Frontier Production Functions," Ph.D. Dissertation, Michigan State University.

Liang, K.Y. and S.L. Zeger (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika* 73.

Mundlak, Y. (1978), "On the Pooling of Time Series and Cross Section Data," *Econometrica* 46.

Papke, L.E. (2005), "The effects of Spending on Test Pass Rates: Evidence from Michigan," *Journal of Public Economics* 89.

Papke 2008 : Papke, L.E. (2008), "The effects of Changes in Michigan's School Finance System," *Public Finance Review* 89.

Papke, L.E. and J.M. Wooldridge (1993), "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics* 89.

Papke, L.E. and J.M. Wooldridge (2008), "Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates," *Journal of Econometrics* 145.

Rivers, D. and Q.H. Vuong (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39.

Rosenzweig, M. R. and Wolpin, K. I. (1980), "Life-Cycle Labor Supply and Fertility: Causal Inferences From Household Models," *Journal of Political Economy* 88.

Terza, J.V. (1998), "Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects," *Journal of Econometrics* 84.

Wooldridge, J.M. (1991b), "Specification Testing and Quasi-Maximum Likelihood Estimation," *Journal of Econometrics* 48.

Wooldridge, J.M. (2002), Econometric Analysis of Cross Section and Panel Data. MIT press, Cambridge, MA.

Wooldridge, J.M. (2005), "Unobserved Heterogeneity and Estimation of Average Partial Effects," in Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. D.W.K. Andrews and J.H. Stock,(Eds.), Cambridge University Press, Cambridge.

Wooldridge, J.M. (2007), "Quasi-LIML Estimation of Nonlinear Models with Endogenous Explanatory Variables," unpublished manuscript, Michigan State University.