Estimating Binary Response Panel Data Models with Selection and Endogeneity: Fertility and Women's Self-Employment

Anastasia Semykina
Department of Economics
Florida State University
Tallahassee, FL 32306-2180
E-mail: asemykina@fsu.edu

This version: January 2016

Estimating Binary Response Panel Data Models with Selection and Endogeneity: Fertility and Women's Self-Employment

Abstract

The existing studies of women's self-employment often focus on the impact of children. Identifying the causal effects of fertility on women's self-employment is complicated because of sample selection and endogeneity of fertility decisions. This paper presents a new estimation approach that considers women's labor force participation, self-employment, and fertility outcomes jointly. The model is estimated using data from the National Longitudinal Survey of Youth 1979 (NLSY79), years 1982-2006. The results indicate that ignoring self-selection into employment and endogeneity of the children variable leads to underestimating the effect of young children. Once both sources of biases are accounted for, the estimated effect of young children on the probability of self-employment more than doubles when compared to the result obtained from a single equation estimation.

JEL Classifications: C33, C34, C35, J13, J22

Keywords: binary response, endogeneity, fertility, sample selection, self-employment

1 Introduction

A rapid growth in female labor force participation during the second half of the last century was accompanied by a substantial increase in self-employment among women. Beginning the mid 1990s, each year about 7% of all female nonagricultural workers in the U.S. (roughly five million women) choose to be self-employed, with numbers being consistently stable over time (Roche, 2014). Understanding the motivation behind women's self-employment choices can aid in developing policies that would help promote economic success among women. However, analyzing the determinants of female self-employment is complicated because of women's extensive involvement in home production and necessity to balance career with family obligations.

While the existing studies of men's employment choices focus primarily on either the entrepreneurial nature of self-employment or its role as a more accessible type of work for those unable to find jobs in the wage sector (Blanchflower et al. 1998, 2001, Carrasco 1999, Earle and Sakova 2000, Evans and Leighton 1989, Koellinger et al. 2012), self-employment among women has been tightly linked to fertility decisions (Boden 1999, Connelly 1992, Constant 2006, Hundley 2000, Lombard 2001, Wellington 2006). The challenge of identifying causal effects of fertility on women's self-employment originates from the presence of the unobserved time-constant and time-varying factors that influence both outcomes. The causality analysis is further complicated by relatively low labor force participation among women, whose decisions to become economically active are often linked to family planning and availability of employment options with flexible work arrangements.

This paper presents a new estimation approach that considers women's labor force participation, self-employment, and fertility outcomes jointly. The proposed estimation strategy extends the parametric approach of Meng and Schmidt (1985) and Semykina and

Wooldridge (2015)[1] to the case of binary endogenous explanatory variables and accounts for three sources of biases: non-random self-selection into the labor force, interrelated nature of fertility and self-employment decisions, and unobserved individual-specific heterogeneity that affects women's choices. Thus, the employed methodology helps to obtain more reliable estimates of causal effects of children and other factors on women's self-employment.

Existing studies often report positive effects of children on the probability of self-employment among women (Boden 1999, Connelly 1992, Edwards and Field-Hendrey 2002, Wellington 2006, for example). This finding is typically explained by women's increased demand for flexible work schedules during times when market work has to be combined with child rearing activities (Boden 1999, Constant 2006, Hundley 2000, Lombard 2001). In contrast to wage work, self-employment hours can be relatively easily adjusted to accommodate family responsibilities, which makes it an attractive option to women with children. However, the analysis presented in earlier papers treats children variables as exogenous – an assumption that is likely violated, especially if women's decisions to have more children are related to their ability to run a successful business. This potential endogeneity problem is discussed in detail in Wellington (2006), but, to the best of our knowledge, it has not been previously addressed in the empirical literature.

In most existing studies, the determinants of self-employment are examined using data for working women. If self-selection into employment is nonrandom, the resulting estimates will suffer from a selection bias. Several authors attempted to address this issue in the past. Macpherson (1988) discusses the importance of correcting for self-selection when estimating earnings equations for self-employed married women. Edwards and Field-Hendrey (2002) use a universal logit model, while Allan and Curington (2014), Connelly (1992), and Hundley (2000) perform multinomial logit estimation to account

[1] Meng and Schmidt (1985) consider binary response sample selection models for cross section, while Semykina and Wooldridge (2015) focus on panel data models.

for the fact that employment and self-employment have to be considered together. An unfortunate feature of the multinomial logit model is the underlying independence of irrelevant alternatives assumption, which rarely holds in practice. The universal logit model overcomes this problem, but addressing endogeneity in this and other logit models is very difficult, especially then the endogenous variable is discrete. In this paper, we specify a separate equation for the employment outcome and estimate it jointly with the self-employment equation. One advantage of this approach is that it fits naturally in the existing literature on labor force participation and sample selection. Importantly, it also allows for a feasible way to address endogeneity of fertility choices in both self-employment and employment outcomes.

Moreover, with few exceptions, the existing analysis of female self-employment is performed using either single-year cross-section data or multiple years of pooled independent cross-sections. A limited longitudinal study was done by Wellington (2006), who used selected years from two longitudinal surveys to examine women's self-employment in each year and in the pooled sample. In her longitudinal exploration, Wellington estimates the effect of the change in the number of children on the probability of entering self-employment and finds no significant effects. Schiller and Crewson (1997) employ longitudinal data to construct cumulative measures of male and female self-employment, although they do not consider the impact of children. Similar to these studies, the present paper employs a large panel data set. However, the implemented estimation utilizes all available years of data and accounts for the presence of unobserved time-constant factors using the techniques suggested in the previous panel data literature.

The data employed in this study comes from the National Longitudinal Survey of Youth 1979 (NLSY79), years 1982-2006. The sample is limited to white women, who comprise the largest self-employed group among all female workers. The results of the partial maximum likelihood estimation indicate that ignoring sample selection and endogeneity leads to the substantial underestimation of the effect of fertility. Once both the endogeneity of fertility decisions and non-randomness in the employment status are accounted for, the estimated effect of young children on the probability of self-employment more than doubles when compared to the result obtained from a single equation estimation.

The rest of the paper is organized as follows. Section 2 presents the econometric model of women's self-employment, while Section 3 summarizes the estimation approach. Data are described in Section 4. Main results and robustness checks are discussed in Sections 5 and 6, respectively. Concluding remarks are provided in Section 7.

2 Econometric Model of Women's Self-Employment

The model of self-employment presented here is similar in spirit to a model of occupational choice. Specifically, a woman becomes self-employed if her utility from self-employment exceeds her utility from working in the wage and salary sector. Modeling the utility difference as a linear function of observed and unobserved factors, the self-employment outcome for woman i in year t can be written as

$$y_{it} = 1[\mathbf{x}_{it}\boldsymbol{\beta} + \gamma d_{it} + a_i + u_{it} > 0], \tag{1}$$

where $1[\cdot]$ is an indicator function equal to one if the expression in brackets is true and zero otherwise, $\mathbf{x}_{it}$ is a vector of observed characteristics (e.g. age, education, income of the spouse), a_i contains time-constant unobserved variables, and u_{it} is an idiosyncratic error that includes unobserved time-varying factors. The key variable of interest is the young children indicator, d_{it}, which equals one if the woman has children ages 0-5.[2] Our

[2] This definition of d_{it} mimics the categorization (young versus older children) used in the majority of the literature on women's self-employment. In Section 6, we consider an alternative specification, where

main focus is on estimating the effect of young children because children's influence on women's labor market choices is typically stronger when children are small (Browning 1992, Carrasco 2001), and because endogeneity concerns are often related to women's decisions to have more children.

As mentioned earlier, self-employment may be preferred by women with children because of the flexibility offered by this type of employment. Hundley (2000) notes the existence of lower and upper bounds on the acceptable levels of productivity and work hours in the wage sector. Because such restrictions are significantly milder among the self-employed, this type of employment provides women with an opportunity to combine home production with market work (Connelly 1992, Constant 2006, Hundley 2000, Lombard 2001, Wellington 2006). Self-employment is expected to be especially attractive to women with young children due to their heavy involvement in housework and childrearing activities.

Estimation of equation (1) is complicated by the presence of sample selection. Although equation (1) is specified for the entire population of working-age women, the outcome in observed only for working females. Formally, observability of y_{it} depends on the outcome in the employment equation:

$$s_{it} = 1[\mathbf{w}_{it}\boldsymbol{\alpha} + \varphi d_{it} + b_i + \epsilon_{it} > 0], \tag{2}$$

where s_{it} is equal to one if woman i works in year t and is zero otherwise, $\mathbf{w}_{it}$ includes observed characteristics, b_i is a time-constant unobserved effect and ϵ_{it} is an idiosyncratic error. Equation (2) follows from a standard theoretical model of labor supply, where a woman's labor force participation decision arises as a solution to the utility maximization problem subject to time and budget constraints, and both constraints depend on the presence of children (see, for example, Moffitt 1984). Estimation wise, equation (2)

d_{it} is defined as an indicator equal to one if the woman has a newborn.

corresponds to the selection rule in a Type 2 Tobit model of labor supply.

In linear cross section models, the most commonly used method for addressing non-random selection is Heckman's correction (Heckman 1979). When applied to panel data, this estimator is generally inconsistent when combined with fixed effects or first difference estimators, but several alternatives have been proposed (Kyriazidou 1997, Rachina-Barrachina 1999, Semykina and Wooldridge 2010, among others). However, when the dependent variable is binary, these correction techniques are invalid. The selection correction term, which captures the dependence between unobservables in the primary and selection equations in models with (roughly) continuous dependent variables, fails to capture the corresponding dependence in limited dependent variable models. Therefore, the estimation approach adopted in this paper builds on parametric methods for binary response models with sample selection considered in Meng and Schmidt (1985) and Semykina and Wooldridge (2015).[3] We extend the model to the case of binary endogenous covariates, which helps to account for the endogeneity of d_{it} in the self-employment and employment equations.

In the literature on women's labor supply, it has long been recognized that fertility is determined jointly with the labor force participation, and several papers suggested ways to resolve this endogeneity issue (Angrist and Evans 1998, Carrasco 2001, Moffitt 1984, among others). A similar endogeneity problem is likely to arise in the self-employment equation, especially when considering family enlargement decisions. To address the endogeneity of the young children indicator, we specify the model for d_{it} as

$$d_{it} = 1[\mathbf{z}_{it}\boldsymbol{\delta} + c_i + v_{it} > 0], \tag{3}$$

where variable definitions are similar to those in equations (1) and (2). In all three

[3]Semiparametric estimators for binary response models with sample selection were proposed by Escanciano et al. (2014) and Semykina and Wooldridge (2015). However, accommodating binary endogenous variables in both main and selection equations is very hard in semiparametric models.

equations we will assume that $\mathbf{x}_{it}$, $\mathbf{w}_{it}$, and $\mathbf{z}_{it}$ are independent of $(u_{it}, \epsilon_{it}, v_{it})$ conditional on the unobserved effects. The problem of endogeneity arises if the unobserved time-varying and time-constant factors in different equations are correlated.

If unobserved effects are independent of covariates and selection is random or $s_{it} = 1$ for all i and t, equations (1) and (2) can be estimated using a switching regression model or a bivariate probit model (Carrasco 2001, Maddala 1983). A control function approach has been proposed by Terza, Basu, and Rathouz (2008), although it is only applicable in cases of "small" endogeneity (i.e. low correlation between the unobservables in the two equations) and requires imposing a relatively strong conditional independence assumption (Wooldridge 2014, 2015). Because the non-randomness in the employment outcome is very likely, the estimator employed in this paper considers all three equations jointly.

The estimation of model parameters is nontrivial due to the presence of time-constant unobserved effects. Since the model is nonlinear, standard panel data methods, such as fixed effects and first difference, would not produce consistent estimators. A commonly recommended solution that was found to be reasonably successful in empirical applications is to use the linear probability model. Unfortunately, the resulting estimator is generally inconsistent because the conditional probability may fall outside of the $[0, 1]$ interval (Horrace and Oaxaca 2006). This method is even less reliable in a multiple equation setup described above. As mentioned earlier, standard selection correction techniques are not applicable in binary response models even when cross-section data are used. Moreover, implementing the fixed-effects estimation on a selected sample would require specifying the conditional distribution of the error in (1) conditional on the employment outcomes in all time periods, which would result is a much more complex selection correction.[4] Therefore, instead, the literature suggests making assumptions about the distribution of the unobserved effect conditional on observed covariates (Chamberlain 1981, Carrasco

[4]For the discussion of this issue in the context of linear panel data models see, for example, Semykina and Wooldridge (2010).

2001, Semykina and Wooldridge 2015, among others). This is the approach adopted in the present paper. Specifically, we use Mundlak (1978) and Chamberlain (1981) approach to model individual time-constant effects as linear functions of exogenous variables:

$$\begin{aligned} a_i &= \mathbf{p}_i \boldsymbol{\xi}_a + \bar{\mathbf{q}}_i \boldsymbol{\psi}_a + r_{ai}, \\ b_i &= \mathbf{p}_i \boldsymbol{\xi}_b + \bar{\mathbf{q}}_i \boldsymbol{\psi}_b + r_{bi}, \\ c_i &= \mathbf{p}_i \boldsymbol{\xi}_c + \bar{\mathbf{q}}_i \boldsymbol{\psi}_c + r_{ci}, \end{aligned} \tag{4}$$

where r_{ai}, r_{bi}, and r_{ci} are unobserved effects that are assumed to be independent of $(\mathbf{p}_i, \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{z}_{it})$. Vector $\mathbf{p}_i$ includes measures of cognitive ability and personality – time-constant characteristics that have been identified in the previous literature as important determinants of self-employment (Evans and Leighton 1989, Schiller and Crewson 1997, Verheul et al. 2012). Although true individual ability and personality are not observed, the available measures ($\mathbf{p}_i$) can often serve as reasonable proxies. To make the independence assumption more reliable, these are also included in equations for b_i and c_i. Vector $\bar{\mathbf{q}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{q}_{it}$ consists of the individual time means of strictly exogenous variables that are likely correlated with time-constant unobservables (other than those in $\mathbf{p}_i$).

Upon substitution, we obtain

$$\begin{aligned} y_{it} &= 1[\mathbf{x}_{it}\boldsymbol{\beta} + \gamma d_{it} + \mathbf{p}_i \xi_a + \bar{\mathbf{q}}_i \boldsymbol{\psi}_a + r_{ai} + u_{it} > 0], \\ s_{it} &= 1[\mathbf{w}_{it}\boldsymbol{\alpha} + \varphi d_{it} + \mathbf{p}_i \boldsymbol{\xi}_b + \bar{\mathbf{q}}_i \boldsymbol{\psi}_b + r_{bi} + \epsilon_{it} > 0], \\ d_{it} &= 1[\mathbf{z}_{it}\boldsymbol{\delta} + \mathbf{p}_i \boldsymbol{\xi}_c + \bar{\mathbf{q}}_i \boldsymbol{\psi}_c + r_{ci} + v_{it} > 0]. \end{aligned} \tag{5}$$

Model parameters can be estimated using the partial maximum likelihood estimator (partial MLE) that is described in detail in the following subsection. This approach allows both time-constant unobserved effects (r_{ai}, r_{bi}, r_{ci}) and idiosyncratic errors (u_{it}, ϵ_{it}, v_{it})

to be correlated across equations and, therefore, accounts for both non-random selection and endogeneity.

When estimating equations (1), (2), and (3) jointly by partial MLE it is not necessary to impose exclusion restrictions. The model can be estimated even if the covariates in all three equations are the same (Wooldridge, 2010). However, the results are more reliable when $\mathbf{z}_{it}$ includes at least one variable that is not in $\mathbf{x}_{it}$ and $\mathbf{w}_{it}$. In identifying such a variable we rely on the previous studies of women's labor supply and employ the instrument proposed by Angrist and Evans (1998). Using the information on children's gender we construct an indicator equal to one if the first two children have the same gender and include it among the covariates in (3). As an exclusion restriction in the employment equation, we assume that employment opportunities influence the employment outcome, but have no direct effect on self-employment and fertility choices. Therefore, the region unemployment rate and region indicators are included in $\mathbf{w}_{it}$, but not in $\mathbf{x}_{it}$ or $\mathbf{z}_{it}$. Later, robustness checks are performed to study the sensitivity of estimation results to the imposed exclusion restrictions.

3 Estimation

To estimate the model in (5), we make parametric assumptions about the error distribution in a given period t. Specifically, let $\mathbf{q}_i = (\mathbf{q}_{i1}, \ldots, \mathbf{q}_{iT})$ and define composite errors, $\eta_{it} = r_{ai} + u_{it}$, $e_{it} = r_{bi} + \epsilon_{it}$, $\zeta_{it} = r_{ci} + v_{it}$, where

$$\eta_{it}, e_{it}, \zeta_{it} | \mathbf{z}_{it}, \mathbf{x}_{it}, \mathbf{w}_{it}, \mathbf{q}_i, \mathbf{p}_i \sim \eta_{it}, e_{it}, \zeta_{it}, \tag{6}$$

$$\begin{pmatrix} \eta_{it} \\ e_{it} \\ \zeta_{it} \end{pmatrix} \sim Normal \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & \\ \rho_{12} & 1 & \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \right), \quad t = 1, \ldots, T. \tag{7}$$

Note that if $\rho_{12} = \rho_{13} = \rho_{23} = 0$, then consistent estimators of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta}$ can be obtained by estimating each equation separately. If $\rho_{12} = 0$ (i.e. self-selection is nonexistent or random), but $\rho_{13} \neq 0$, then endogeneity should be addressed by estimating (1) and (3) jointly. Finally, if $\rho_{12} \neq 0$ and $\rho_{13} = \rho_{23} = 0$, so that d_{it} is exogenous but there is self-selection into employment, then joint estimation of (1) and (2) is required for consistency.

We construct the likelihood function under the assumption that all correlation coefficients may potentially be different from zero. Specifically, for a working woman there are four possible outcomes. She may be self-employed with young kids, self-employed without young kids, or wage-employed with or without young children. Correspondingly, for nonworking women there are two possibilities: they may have little children or not. Therefore, the likelihood function for a woman i in year t is

$$\begin{aligned} l_{it} &= P_{111,it}^{s_{it} y_{it} d_{it}} \times P_{110,it}^{s_{it} y_{it} (1-d_{it})} \times P_{101,it}^{s_{it}(1-y_{it}) d_{it}} \times P_{100,it}^{s_{it}(1-y_{it})(1-d_{it})} \\ &\times \; P_{01,it}^{(1-s_{it}) d_{it}} \times P_{00,it}^{(1-s_{it})(1-d_{it})}, \end{aligned} \tag{8}$$

where $P_{111,it} \equiv P(s_{it} = 1, y_{it} = 1, d_{it} = 1)$, $P_{110,it} \equiv P(s_{it} = 1, y_{it} = 1, d_{it} = 0)$, $P_{101,it} \equiv$

$P(s_{it}=1, y_{it}=0, d_{it}=1)$, etc., and

$$\begin{aligned}
P_{111,it} &= \int_{-\infty}^{\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{\tilde{\mathbf{x}}_{it}\tilde{\boldsymbol{\beta}}} \int_{-\infty}^{\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_3(\varepsilon, \mu, \nu; \rho_{12}, \rho_{13}, \rho_{23})\ d\varepsilon\ d\mu\ d\nu, \\
P_{110,it} &= \int_{-\infty}^{\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{\tilde{\mathbf{x}}_{it}\tilde{\boldsymbol{\beta}}} \int_{-\infty}^{-\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_3(\varepsilon, \mu, \nu; \rho_{12}, -\rho_{13}, -\rho_{23})\ d\varepsilon\ d\mu\ d\nu, \\
P_{101,it} &= \int_{-\infty}^{\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{-\tilde{\mathbf{x}}_{it}\tilde{\boldsymbol{\beta}}} \int_{-\infty}^{\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_3(\varepsilon, \mu, \nu; -\rho_{12}, \rho_{13}, -\rho_{23})\ d\varepsilon\ d\mu\ d\nu, \\
P_{100,it} &= \int_{-\infty}^{\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{-\tilde{\mathbf{x}}_{it}\tilde{\boldsymbol{\beta}}} \int_{-\infty}^{-\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_3(\varepsilon, \mu, \nu; -\rho_{12}, -\rho_{13}, \rho_{23})\ d\varepsilon\ d\mu\ d\nu, \\
P_{01,it} &= \int_{-\infty}^{-\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_2(\varepsilon, \nu; -\rho_{13})\ d\varepsilon\ d\nu, \\
P_{00,it} &= \int_{-\infty}^{-\tilde{\mathbf{w}}_{it}\tilde{\boldsymbol{\alpha}}} \int_{-\infty}^{-\tilde{\mathbf{z}}_{it}\tilde{\boldsymbol{\delta}}} \phi_2(\varepsilon, \nu; \rho_{13})\ d\varepsilon\ d\nu,
\end{aligned} \tag{9}$$

$\phi_2(.)$ and $\phi_3(.)$ denote the probability density functions of the bivariate and trivariate standard normal distributions, respectively; $\tilde{\mathbf{x}}_{it} \equiv (\mathbf{x}_{it}, d_{it}, \mathbf{p}_i, \bar{\mathbf{q}}_i)$, $\tilde{\mathbf{w}}_{it} \equiv (\mathbf{w}_{it}, d_{it}, \mathbf{p}_i, \bar{\mathbf{q}}_i)$, $\tilde{\mathbf{z}}_{it} \equiv (\mathbf{z}_{it}, \mathbf{p}_i, \bar{\mathbf{q}}_i)$, $\tilde{\boldsymbol{\beta}} \equiv (\boldsymbol{\beta}', \gamma, \boldsymbol{\xi}_a', \boldsymbol{\psi}_a')'$, $\tilde{\boldsymbol{\alpha}} \equiv (\boldsymbol{\alpha}', \varphi, \boldsymbol{\xi}_b', \boldsymbol{\psi}_b')'$, and $\tilde{\boldsymbol{\delta}} \equiv (\boldsymbol{\delta}', \boldsymbol{\xi}_c', \boldsymbol{\psi}_c')'$.

The estimation is implemented using the maximum simulated likelihood estimator (Cappellari and Jenkins 2006, Train 2003), which is less computationally demanding than numerical integration methods. Because the conditional time dependence in $\{y_{it}, s_{it}, d_{it}\}_{t=1}^T$ is likely present, we obtain standard errors that account for serial dependence, which may be due to both the unobserved time-constant factors that are part of the composite error and serial correlation in idiosyncratic shocks. In this paper, the standard errors are obtained using panel bootstrap.

In binary response panel data models the parameters of interest are typically average partial effects (APEs), which are the partial effects averaged over the distribution of the unobserved effect. For example, the APE of small children on the probability of self-employment is obtained by considering changes in $\mathrm{E}_a[P(y_{it}=1|\mathbf{x}_{it}, d_{it}, a)]$ when d_{it} changes from 0 to 1. Using theoretical results from Papke and Wooldridge (2007), this

APE can be consistently estimated as

$$\begin{aligned} & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\Phi_{it}^{1}-\Phi_{it}^{0}), \qquad (10) \\ \Phi_{it}^{1} &= \Phi(\mathbf{x}_{it}\hat{\boldsymbol{\beta}}+\hat{\gamma}+\mathbf{p}_i\hat{\boldsymbol{\xi}}_a+\bar{\mathbf{q}}_i\hat{\boldsymbol{\psi}}_a), \\ \Phi_{it}^{0} &= \Phi(\mathbf{x}_{it}\hat{\boldsymbol{\beta}}+\mathbf{p}_i\hat{\boldsymbol{\xi}}_a+\bar{\mathbf{q}}_i\hat{\boldsymbol{\psi}}_a), \end{aligned}$$

where $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$, $\hat{\boldsymbol{\xi}}_a$, and $\hat{\boldsymbol{\psi}}_a$ are the estimated model parameters, and $\Phi(\cdot)$ denotes the cumulative distribution function of the univariate standard normal distribution. The APEs of other variables on self-employment, as well as APEs on probabilities of employment and having young children are computed similarly.

4 Data

The model presented in Section 2 is estimated using data from the National Longitudinal Survey of Youth 1979 (NLSY79), years 1982-2006. The data set is a nationally representative sample of individuals who were 14 to 22 years old in 1979. Surveys were conducted annually up until 1994 and every other year afterwards. The age of the oldest respondent in 2006 was 49. The sample employed in the analysis is limited to white women ages 22 and older, who are not enrolled in school and do not work in agriculture. Women serving in the military are excluded. The final sample consists of 33,211 person-year observations, where 28,234 observations correspond to the cases when women worked.

Employment and self-employment indicators were created using information on the current or most recent job. As is standard in the literature, the self-employed include both incorporated and unincorporated self-employed.[5] An indicator for young children, d_{it}, was constructed using information on the age of currently living children. An indicator for

[5]Separate analyses of incorporated and unincorporated self-employed are practically infeasible due to small number of observations in each group and modeling complications.

having older children (ages 6-17) is also included and treated as predetermined. Other explanatory variables include age, age squared, educational attainment indicators, an indicator for urban locales, marital status indicator, and income of the spouse in thousands of dollars. Income data were deflated to 2000 dollars using the consumer price index. As mentioned earlier, region indicators and the unemployment rate are included in the employment equation, while the same sex indicator (equal to one if the first two children are of the same gender) is included in the young children equation.

As a measure of cognitive ability we use the Armed Forces Qualification Test (AFQT) score from 1979. We also use two personality measures: self-esteem and locus of control. The self-esteem scale was developed by Rosenberg and measures the degree of approval or disapproval toward oneself (Rosenberg, 1965). In the estimation, we use results from a test that was administered in 1980. The locus of control measure was proposed by Rotter and describes the extent to which the person believes own life outcomes being determined by own actions rather than fate or luck (Rotter, 1966). In the NLSY79, the Rotter scale was administered in 1979 and is negative for individuals who believe that they control own lives. For interpretational convenience, cognitive ability and personality measures were standardized to have zero mean and unit variance.

The vector of individual time means, $\bar{q}_i$, includes marital status and income of the spouse because these two variables vary over time and are likely correlated with time-constant unobservables, other than those captured by p_i. In the present context, the time means can be interpreted as measures of the financial stability and solidity of marriage, which can also be viewed as proxies for unobserved time-invariant characteristics (e.g. cultural factors, family background) that can make the marriage more or less successful. The other variables (e.g age, education, presence of older children, region of residence) are assumed to be related to the unobserved time-invariant effects only through their correlation with ability, personality traits, and time-means of the marital status and spouse's income.

Sample summary statistics for the employed, self-employed, and nonworking women are presented in Table 1. As seen in the Table, self-employed women in the sample tend to be slightly older and well educated, while women who do not work comprise the least educated group. Self-employed and nonworking women are much more likely to be married and have children than women who work in the wage sector. Spouse's income is the highest among nonworking females and lowest among the wage-employed. Most self-employed women reside in cities, while nonworking females in the sample are more often located in rural areas. Wage-employed and nonworking women are slightly more likely to reside in the North Central region and in the South, while self-employed respondents are more commonly found in the West. Finally, self-employed women tend to have slightly higher AFQT scores, a more internal locus of control, and higher self-esteem. The average values of the ability and personality scores are the lowest for nonworking women.

Table 2 displays the self-employment rates by age for the entire sample, as well as the percent of wage-employed, self-employed, and nonworking women in different age groups who have young and older children. All statistics are obtained using pooled data. It is evident that more women become self-employed as they age (column 1 in Table 2). However, there does not appear to be much variation after the age of 30. As seen in columns (2) through (7) in Table 2, nonworking and self-employed women are most likely to have small children in their twenties, while among wage-employed women the peak is in the late twenties and early thirties. As expected, the percent of women with young children is very small among the oldest group. That is, the data cover both the times of the highest family demands and years when home responsibilities for women are somewhat reduced.

5 Results

In this section, we discuss the results from estimating the self-employment equation presented in Section 2. To assess the importance of correcting for endogeneity and selection, we first use the sample of working women and estimate the self-employment equation separately. Then, estimation is done for two-equation models that correct for either nonrandom selection or endogeneity of the young children variable. Subsequently, the full three-equation model is estimated. In all equations, year indicators are included to account for common time-specific shocks. Standard errors were obtained using panel bootstrap with 100 replications and are robust to serial correlation.

The estimated partial effects from a single-equation probit regression are displayed in column (1) of Table 3. The partial effect of age is estimated as a predicated change in the probability of self-employment when age increases from 31.2 (the average age in the sample) to 32.2 years. The effect is positive (about 4.6 percentage points increase) and highly statistically significant. Moreover, based on the estimated coefficients, the relationship between age and self-employment is positive almost for the entire age range observed in the data. Although the age profile is slightly concave (the coefficient on age is positive, and the coefficient on age squared is negative), the turning point is estimated to be at about 46.7 years, which is close to the age of the oldest woman in the sample (49 years).

Furthermore, it appears that better educated women are slightly less likely to be self-employed. However, the estimated effects of education indicators are individually and jointly insignificant (the Wald test statistic for the joint significance test is 2.07, p-value = 0.5589), indicating no systematic relationship between education and self-employment. In contrast, the likelihood of self-employment is strongly positively related to the income of the spouse and is higher among married women. Although personality and cognitive ability have expected effects – the probability of self-employment is positively related to AFQT and self-esteem scores and is slightly higher among women with a more internal locus of control – the estimates are not statistically significant.

Similar to findings reported in previous studies, the incidence of self-employment in a given period is positively associated with the presence of children. The probability of being self-employed is estimated to be approximately 4.4 percentage points higher among women who have young children than among women without children. The estimated partial effect of having older children is about 50 % smaller (2.22 percentage points), which is as expected. While caring for children ages 6-17 often requires substantial amount of time, these demands tend to be considerably lower than in the case of young children. The partial effects of both young and older children are highly statistically significant.

Columns (2) and (3) in Table 3 present results from estimating the employment and self-employment equations jointly. As seen in column (2), the estimated average partial effect of young children increases to approximately 5.9 percentage points and remains highly significant. The effect of older children also increases slightly. The estimated impact of spouse's income become less statistically significant, while the estimated partial effects of other variables change very little.

In the employment equation (column 3, Table 3), the estimated APEs on the probability of working are largely consistent with our expectations. The estimated effect of age is negative, which is similar to the results presented in other studies. Moreover, there is a strong positive relationship between woman's probability of employment and education. For example, having more than 15 years of schooling is associated with an about 17 percentage points higher chance of employment as compared with a woman who has less than 12 years of education. The probability of working is also greater among females who have higher AFQT and self-esteem scores and reside in urban locales. The income of the spouse is strongly negatively related to employment. The APE of the region unemployment rate is also negative, although only marginally significant.

As expected, women with children are less likely to work, and the estimated impact of young children is notably larger than that of older children. Specifically, the probability of working among women with children ages 0-5 is about 17 percentage points lower than among females without children. The estimated partial effect of older children is roughly a quarter of that number.

When the self-employment equation is estimated jointly with the young children equation (last two columns in Table 3), the partial effect of young children on self-employment increases to about 10.5 percentage points and is highly significant. In other words, it appears what ignoring the endogeneity of fertility decisions leads to underestimating the effect of young children by a large margin. The estimated partial effects of the marital status and spouse's income also change (they decrease by approximately 50% and become less significant), which is probably not surprising because both of these variables are substantially correlated with the young children indicator. The estimated effect of older kids is rather similar to that in columns (1) and (2). The APEs of other variables are also nearly unchanged.

In the young children equation (last column in Table 3), the most important determinants are education, age, and family variables. A one year increase in age from the sample mean (about 31.2 years) is associated with a 1 percentage point reduction in the predicted probability of having young children. Indeed, the estimates suggest that the age profile is slightly concave (the coefficient on the linear age term is positive, while the coefficient on the quadratic term is negative), with the estimated turning point being 27.3 years old. Consistent with our expectations, the effect of education is negative. The estimated difference between the most educated (more than 15 years of schooling) and the least educated (less than 12 years of schooling) women is more than 10 percentage points. Moreover, the likelihood of having young children is higher among married women and is positively related to the income of the spouse, which is a expected. Neither cognitive ability nor personality appear to matter, while the effect of the indicator for the first two children having the same gender is positive and highly significant.

Table 4 reports the results from estimating all three equations together. The estimated APEs of education indicators on the probability of self-employment in a given year t are still insignificant, while the effect of age increases in magnitude and remains significant. The estimated effect of spouse's income loses some of its significance, while the marital status is not significant even at the 10% level. The directions of these changes coincide with those in Table 3. Importantly, the estimated partial effect of young children on self-employment in Table 4 is almost three times the estimate obtained from a single equation model (column 1, 3). This is consistent with previously presented results and provides evidence suggesting that ignoring nonrandom sample selection and endogeneity leads to underestimating the effect of d_{it}.

In the employment equation (column 2, Table 4), the estimated partial effects of age and marital status substantially increase in magnitude compared to the two equation model (column 3, Table 3). The partial effect of young children nearly doubles in magnitude, which is similar to the findings of Carrasco (2001), who corrects for the endogeneity of the young children variable in the women's labor force participation equation. In Table 4, the likelihood of working among women with young children is approximately 33.5 percentage higher than among women without children. In contrast, the magnitude of the estimated partial effect of older children decreases slightly. The effects of other variables are similar to those in Table 3, column (3).

When the young children equation is estimated jointly with the employment and self-employment equations (column 3 in Table 4), the estimated partial effects are rather similar to those presented in Table 3, column (5). This is as expected because the young children model is a reduced-form equation, where accounting for error correlations across equations should not affect the estimates in a significant way.

To better understand the causes of the observed differences between estimated APEs in Tables 3 and 4, it is useful to look at the estimated correlation coefficients for the errors in the three equations. Coefficient estimates obtained from estimating the full three equation model are presented in Table 5. As seen in the Table, the correlation coefficient between errors in the young children and self-employment equations (ρ_{13}) is negative and large in magnitude. It is also highly statistically significant. This finding suggests that women without young children are more likely to be self-employed due to other, unobserved factors. For example, these women may have stronger career aspirations and desire for independence (being one's own boss). The correlation may also be due to time-specific shocks. For example, a temporary reduction in wage employment opportunities due to a sudden worsening of economic conditions may motivate women to start a new business or remain self-employed for a longer period of time (a "push" or "refugee" effect discussed in the literature on male self-employment). The same negative economic shock may motive women to postpone having more children until the economy improves. Using the results for mixing normal distributions, the negative correlation between errors in the two equations can be interpreted in the context of an omitted variable problem, where one omits an important variable that is positively related to the probability of having young children and has a negative partial effect on self-employment. Therefore, ignoring the endogeneity of the young children indicator leads to a negative bias and underestimating the effect of d_{it} on the probability of being self-employed.

The correlation between the errors in the self-employment and employment equations is also negative (ρ_{12} in Table 5), which should lead to underestimating the APE of young children in the model of self-employment that ignores selection into working. However, $\mathrm{Corr}(\eta_{it}, e_{it})$ is highly statistically insignificant. Therefore, it is not surprising that accounting for sample selection has a minor impact on the results. This suggests that if a nonworking woman were to become employed, her probability of holding a job in the wage sector versus running her own business would be similar to that for a comparable currently employed woman. In other words, in these data the bias appears to be mostly due to nonrandomness in fertility decisions.

Another interesting finding is that the correlation between errors in the employment and young children equations is positive and highly statistically significant (ρ_{23} in Table 5). This indicates the presence of unobserved factors that increase the probability that a woman will have small children in a given year and also increase the likelihood that the woman will work. This may happen, for example, if an actual or anticipated improvement in the overall economy motivates women to have more children and also increases their chances of being employed. Using the argument similar to the one presented above, the positive correlation can be interpreted as a case of omitting a relevant variable that is positively associated with the young children indicator and has a positive partial effect of the probability of working. Consequently, the failure to account for endogeneity of young children in the employment equation should lead to a positive bias. Because the estimated effect of young children on employment is negative, the bias implies underestimating the impact of d_{it} in absolute value, which is exactly what we find.

6 Robustness Checks

As discussed in Sections 2 and 4, the performed estimation relies on several exclusion restrictions. Specifically, the indicator for South location and regional unemployment rate are assumed to have an impact on the employment outcome, but are excluded from the self-employment equation. Because the maximum likelihood method does not require imposing exclusion restrictions, the sensitivity of results to these assumptions can be examined. To do so, the full three-equation model was estimated with the South indicator and unemployment rate entering both employment and self-employment equations. Results are presented in Table 6, column (1). Only the estimated average partial effects on the probability of self-employment are reported.[6]

As seen in Table 6, residing in the South has no systematic effect on the probability of self-employment. However, the estimated effect of the local unemployment rate is positive and significant at the 5% level (the coefficient on the unemployment rate in the employment equation is negative and significant at the 5% level). This result indicates that the "push" effect found in the studies of male self-employment (an increased motivation for entering self-employment at times of an economic downturn when finding a job in the wage and salary sector becomes hard) may be present for women, also. Importantly, the partial effect of young children changes little when compared to the result displayed in column (1), Table 4. The estimate is slightly increased and is still highly statistically significant. This suggests that the estimates are not very sensitive to the exclusion restrictions in the employment equation.

Another set of robustness checks considers the sensitivity of results to using the same gender indicator in the young children equation. Column (2) in Table 6 demonstrates how the estimates change when the number of siblings that the woman had in 1979 (instead of the own children's same sex indicator) is included in the model for d_{it}. The motivation for using this alternative variable comes from an anecdotal evidence that individuals who have more siblings tend to have more children of their own. Indeed, the number of siblings has a highly significant positive effect on $\mathrm{P}(d_{it} = 1)$ in the data.[7] As for the partial effect of young children on self-employment, it is still large and positive, which is similar to Table 4. Finally, when neither the number of siblings nor the same gender indicator is included (column 3 in Table 6), the estimated APE of young children does not change

[6]The estimates in the employment and young children equations are available from the author upon request.

[7]Detailed results are not reported, but are available from the author upon request.

much, suggesting that imposing exclusion restrictions in the young children equation has a minor impact on the estimates.

Finally, we consider an alternative definition of the young children indicator. Although the majority of the literature defines young kids as those ages 0-5, it would also be useful to see the effect of new births (i.e. consider changes at the margin). Hence, we redefine d_{it} to equal one if the youngest child is less then one year old, and set it to zero otherwise. An indicator for children ages 1-5 is included as an additional covariate. The results are reported in Table 7. Similar to previous findings, the estimated average partial effect of very young children is positive and significant in a single-equation probit regression (column 1). When moving from one-equation to two- and three-equation models (columns 2 through 4), the magnitude of the effect increases substantially. However, the estimate is insignificant in the models that account for the endogeneity of fertility choices (columns 3 and 4). After looking closely at the data, it appears that the lack of significance may be due to heterogeneity among mothers with very young children. Specifically, among the women in the sample who have other young children (ages 1-5) at the time when a new child is born, the self-employment rate is rather high (roughly 14%). However, among new mothers with no other young children the self-employment rate is only about 6%, which is even less than the self-employment rate among women without children (approximately 7.7%). That is, women who have a baby for the first time seem to be much less confident in own ability to run a business and bear childcare responsibilities simultaneously. This heterogeneity appears to cause the effect of very young children to be less systematic and, hence, less precise.

7 Concluding Remarks

In this paper, we use NLSY79 data to investigate the impact of children on self-employment outcomes among women. The estimation is based on a three equation model that accommodates the interrelated nature of fertility, employment, and self-employment decisions. We find that in these data, the correlation between unobservables that determine employment and self-employment outcomes is not very systematic, so that accounting for self-selection into working has a minor impact on the estimated effect of young children on self-employment. However, when considering employment and self-employment outcomes jointly with fertility, the magnitude of the estimated effects of young kids increases substantially. Results from the full model estimation suggest that the probability of self-employment is approximately 12 percentage points higher among women with young children, while the probability of employment is about 33 percentage points lower than among women without children. These findings suggest that accounting for the endogeneity of fertility in the self-employment and employment equations matters.

Given that children play such a significant role in women's labor market decisions, much more needs to be done to better understand women's employment and self-employment choices when they have kids. For example, an important question is whether a motherhood wage penalty exists for self-employed women, and how it compares to the penalty experienced by mothers employed in the wage sector. Such analysis would require a careful consideration of several factors: self-selection into the labor force, unobserved heterogeneity, and endogeneity of fertility and self-employment choices. The approach presented in this paper can serve as a starting point in developing the appropriate estimation strategy.

References

Allan, David W., and William P. Curington. 2014. "The Self-Employment of Men and Women: What are their Motivations?" *Journal of Labor Research* 35: 143-61.

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *American Economic Review* 88(3): 450-77.

Blanchflower, David G. and Andrew J. Oswald. 1998. "What Makes an Entrepreneur?" *Journal of Labor Economics* 16(1): 26-60.

Blanchflower, David G., Andrew J. Oswald., and Alois Stutzer. 2001. "Latent Entrepreneurship Across Nations." *European Economic Review* 45: 680-91.

Boden, Richard J. 1999. "Flexible Working Hours, Family Responsibilities, and Female Self-Employment." *American Journal of Economics and Sociology* 58(1): 71-83.

Browning, Martin. 1992. "Children and Household Economic Behavior." *Journal of Economic Literature* 30: 1434-75.

Cappellari, Lorenzo, and Stephen Jenkins. 2006. "Calculation of Multivariate Normal Probabilities by Simulation, with Applications to Maximum Simulated Likelihood Estimation." *Stata Journal* 6(2): 156-89.

Carrasco, Raquel. 1999. "Transitions to and from Self-Employment in Spain: An Empirical Analysis." *Oxford Bulletin of Economics and Statistics* 61(3): 315-41.

Carrasco, Raquel. 2001. "Binary Choice With Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation." *Journal of Business and Economic Statistics* 19(4): 385-94.

Connelly, Rachel. 1992. "Self-Employment and Providing Child Care." *Demography* 29(1): 17-29.

Constant, Amelie. 2006. "Female Proclivity to the World of Business." *Kyklos* 59(4): 465-80.

Earle, John S., and Zuzana Sakova. 2000. "Business Start-ups or Disguised Unemployment? Evidence on the Character of Self-Employment from Transition Economies." *Labour Economics* 7: 575-601.

Edwards, Linda N., and Elizabeth Field-Hendrey. 2002. "Home-Based Work and Women's Labor Force Decisions." *Journal of Labor Economics*, 20(1): 170-200.

Escanciano, Carlos E., David T. Jacho-Chávez, and Arthur Lewbel. 2014. "Uniform Convergence of Weighted Sums of Non and Semiparametric Residuals for Estimation and Testing." *Journal of Econometrics*, 178: 426-43.

Evans, David S., and Linda S. Leighton. 1989. "Some Empirical Aspects of Entrepreneurship." *American Economic Review* 79(3): 519-35.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153-61.

Horrace, William C., and Ronald L. Oaxaca. 2006. "Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model." *Economics Letters* 90(3): 32127.

Hundley, Greg. 2000. "Male/Female Earnings Differences in Self-Employment: The Effects of Marriage, Children, and the Household Division of Labor." *Industrial and Labor Relations Review* 54(1): 95-114.

Koellinger, Philipp D., and A. Roy Thurik. 2012. “Entrepreneurship and the Business Cycle.” *Review of Economics and Statistics* 94(4): 1143-56.

Kyriazidou, Ekaterini. 1997. “Estimation of a Panel Data Sample Selection Model.” *Econometrica* 65: 1335-64.

Lombard, Karen V. 2001. “Female Self-Employment and Demand for Flexible, Nonstandard Work Schedules.” *Economic Inquiry* 39(2): 214-37.

Macpherson, David A. 1988. “Self-Employment and Married Women.” *Economics Letters* 28(3): 281-84.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge U.K.: Cambridge University Press.

Meng, Chun-Lo, and Peter Schmidt. 1985. “On the Cost of Partial Observability in the Bivariate Probit Model.” *International Economic Review* 26(1): 71-85.

Moffitt, Robert. 1984. “Profiles of Fertility, Labor Supply and Wages of Married Women: A Complete Life-Cycle Model.” *Review of Economic Studies* LI: 263-78.

Mundlak, Yair. 1978. “On the Pooling of Time Series and Cross Section Data.” *Econometrica* 46(1): 69-85.

Roche, Kristen 2014 “Female Self-Employment in the United States: an Update to 2012.” *Monthly Labor Review*,

http://www.bls.gov/opub/mlr/2014/article/pdf/female-self-employment-in-the-united-states-an-update-to-2012.pdf

Rochina-Barrachina, M.E. 1999. “A New Estimator for Panel Data Sample Selection Models.” *Annales dEconomie et de Statistique* 55/56: 153-81.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image.* Princeton, NJ: Princeton University Press.

Rotter, J. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs: General and Applied* 80(1): 609.

Semykina, Anastasia, and Jeffrey M. Wooldridge. 2010. "Estimating Panel Data Models in the Presence of Endogeneity and Selection." *Journal of Econometrics* 157: 375-80.

Semykina, Anastasia, and Jeffrey M. Wooldridge. 2015. "Binary Response Panel Data Models with Sample Selection and Self Selection." Working Paper,

http://econpapers.repec.org/paper/fsuwpaper/wp2015_5f05_5f01.htm

Schiller, Bradley R., and Phillip E. Crewson. 1997. "Entrepreneurial Origins: A Longitudinal Inquiry." *Economic Inquiry* 35: 523531.

Terza, Joseph, Anirban Basu, and Paul Rathouz. 2008. "Two- Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling." *Journal of Health Economics* 27(3): 53143.

Train, K. 2003. *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

Verheul, Ingrid, Roy Thurik, Isabel Grilo, and Peter van der Zwan. 2012. "Explaining Preferences and Actual Involvement in Self-Employment: Gender and the Entrepreneurial Personality." *Journal of Economic Psychology* 33: 325-41.

Wellington, Alison J. 2006. "Self-Employment: the New Solution for Balancing Family and Career?" *Labour Economic* 13: 357-86.

Wooldridge, Jeffery M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd edition. Cambridge: MIT Press.

Wooldridge, Jeffery M. 2014. "Quasi-Maximum Likelihood Estimation and Testing for Nonlinear Models with Endogenous Explanatory Variables." *Journal of Econometrics* 182(1): 226-34.

Wooldridge, Jeffery M. 2015. "Control Function Methods in Applied Econometrics." *Journal of Human Resources* 50(2): 420-45.

Table 1: Descriptive statistics, NLSY79, white women, 1982-2006.

	Wage-employed	Self-employed	Not working
Age	31.12	32.58	31.30
	(6.64)	(6.38)	(6.54)
Less than 12 years of schooling (%)	7.78	7.24	22.38
12 years of schooling (%)	44.63	44.81	48.60
13 to 15 years of schooling (%)	21.08	21.61	14.82
More than 15 years of schooling (%)	26.51	26.35	14.21
Married (%)	58.20	73.35	69.92
Has children ages 0-5 (%)	31.52	47.46	65.17
Has children age 6-17 (%)	36.47	49.00	52.79
Spouse's income/1000	40.59	51.41	56.34
	(34.44)	(44.95)	(53.85)
Urban (%)	71.90	73.15	66.38
Northeast (%)	19.80	18.21	16.90
Northcentral (%)	31.30	29.94	33.20
South (%)	32.99	29.59	32.62
West (%)	15.91	22.26	17.28
AFQT	0.06	0.14	-0.35
	(0.98)	(0.99)	(1.05)
Locus of control	-0.02	-0.12	0.13
	(0.99)	(1.01)	(1.04)
Self-esteem	0.03	0.14	-0.18
	(0.99)	(0.99)	(1.05)
Number of observations	26,442	2,004	5,205

Table 2: Percent self-employed and percent with children by age and work status.

	Full sample	Wage-employed		Self-employed		Not working	
	Self-employed	Has kids ages 0-5	Has kids ages 6-17	Has kids ages 0-5	Has kids ages 6-17	Has kids ages 0-5	Has kids ages 6-17
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Ages 22-24	3.51	30.07	5.62	59.38	12.50	81.53	20.38
Ages 25-29	6.27	39.19	23.88	62.28	36.49	79.98	47.94
Ages 30-34	9.03	40.19	45.38	54.21	54.38	69.24	61.97
Ages 35-39	8.19	26.69	64.94	37.74	68.39	54.50	75.29
Ages 40-44	8.06	9.52	63.04	20.62	59.92	21.56	71.00
Ages 45-49	9.03	1.85	47.86	2.94	57.84	2.30	49.31

Reported numbers are percentages.

Table 3: Estimated average partial effects in one and two equation models.

	1 Eq. Model	2 Eq. Model		2 Eq. Model	
	y Eq.	y Eq.	s Eq.	y Eq.	d Eq.
	(1)	(2)	(3)	(4)	(5)
Less than 12 years of schooling	0.0046	0.0134	-0.1206***	0.0013	0.0491***
	(0.0117)	(0.0209)	(0.0180)	(0.0120)	(0.0166)
13 to 15 years of schooling	-0.0048	-0.0072	0.0286**	-0.0035	-0.0311**
	(0.0092)	(0.0111)	(0.0115)	(0.0098)	(0.0125)
More than 15 years of schooling	-0.0135	-0.0178	0.0514***	-0.0110	-0.0554***
	(0.0094)	(0.0122)	(0.0122)	(0.0099)	(0.0121)
Age	0.0046***	0.0056**	-0.0049***	0.0053***	-0.0103***
	(0.0016)	(0.0023)	(0.0016)	(0.0016)	(0.0013)
Married	0.0219***	0.0243***	-0.0018	0.0120	0.2470***
	(0.0060)	(0.0074)	(0.0083)	(0.0077)	(0.0112)
Has children ages 0-5	0.0437***	0.0589**	-0.1723***	0.1046***	
	(0.0057)	(0.0238)	(0.0082)	(0.0315)	
Has children age 6-17	0.0221***	0.0280***	-0.0462***	0.0205***	
	(0.0062)	(0.0098)	(0.0067)	(0.0063)	
Spouse's income/1000	0.00017***	0.00025*	-0.00076***	0.00014**	0.00076***
	(0.00006)	(0.00013)	(0.00010)	(0.00006)	(0.00016)
Urban	0.0045	0.0037	0.0251***	0.0045	0.0054
	(0.0065)	(0.0079)	(0.0065)	(0.0069)	(0.0080)
AFQT	0.0019	0.0002	0.0308***	0.0015	0.0084
	(0.0049)	(0.0067)	(0.0044)	(0.0050)	(0.0061)
Locus of control	-0.0037	-0.0043	0.0010	-0.0040	0.0001
	(0.0037)	(0.0044)	(0.0050)	(0.0039)	(0.0052)
Self-esteem	0.0043	0.0043	0.0088*	0.0043	-0.0017
	(0.0037)	(0.0044)	(0.0048)	(0.0039)	(0.0049)
South			0.0113		
			(0.0086)		
Regional unemployment rate			-0.0095*		
			(0.0049)		
First two children are of same gender					0.2437***
					(0.0105)
Number of observations	28,446	33,648		28,446	

Standard errors adjusted for serial correlation are in parentheses.
Reference categories: 12 years of schooling, West region.
All equations also include individual time means for marital status and spouse's income.

Table 4: Estimated average partial effects in a full, three equation model.

	3 Equation Model		
	y Eq.	s Eq.	d Eq.
	(1)	(2)	(3)
Less than 12 years of schooling	0.0063	-0.0964***	0.0768***
	(0.0166)	(0.0169)	(0.0163)
13 to 15 years of schooling	-0.0052	0.0234**	-0.0323***
	(0.0112)	(0.0116)	(0.0116)
More than 15 years of schooling	-0.0143	0.0436***	-0.0583***
	(0.0113)	(0.0128)	(0.0124)
Age	0.0063***	-0.0062***	-0.0113***
	(0.0022)	(0.0015)	(0.0016)
Married	0.0115	0.0320***	0.2579***
	(0.0091)	(0.0096)	(0.0114)
Has children ages 0-5	0.1238***	-0.3345***	
	(0.0470)	(0.0248)	
Has children age 6-17	0.0251***	-0.0386***	
	(0.0082)	(0.0063)	
Spouse's income/1000	0.00020**	-0.00064***	0.00094***
	(0.00010)	(0.00009)	(0.00016)
Urban	0.0038	0.0249***	0.0085
	(0.0081)	(0.0063)	(0.0079)
AFQT	0.0006	0.0310***	0.0008
	(0.0063)	(0.0045)	(0.0058)
Locus of control	-0.0045	0.0013	-0.0003
	(0.0046)	(0.0049)	(0.0046)
Self-esteem	0.0045	0.0085*	-0.0035
	(0.0046)	(0.0048)	(0.0050)
South		0.0109	
		(0.0081)	
Regional unemployment rate		-0.0099**	
		(0.0047)	
First two children are of same gender			0.2586***
			(0.0091)
Number of observations		33,648	

Standard errors adjusted for serial correlation are in parentheses.
The reference category is 12 years of schooling.
All equations also include individual time means for marital status and spouse's income.

Table 5: Estimated error correlation coefficients.

	Coefficient
$\rho_{12} \equiv \text{Corr}(\eta_{it}, e_{it})$	-0.2233
	(0.1811)
$\rho_{13} \equiv \text{Corr}(\eta_{it}, \zeta_{it})$	-0.2401***
	(0.0938)
$\rho_{23} \equiv \text{Corr}(e_{it}, \zeta_{it})$	0.4037***
	(0.0547)

Standard errors adjusted for serial correlation are in parentheses.

Table 6: Robustness checks: Estimated average partial effects on the probability of self-employment when exclusion restrictions vary (full, three equation model).

	y Eq. (1)	y Eq. (2)	y Eq. (3)
Less than 12 years of schooling	0.0136	0.0107	0.0033
	(0.0192)	(0.0152)	(0.0165)
13 to 15 years of schooling	-0.0059	-0.0060	-0.0037
	(0.0123)	(0.0111)	(0.0117)
More than 15 years of schooling	-0.0157	-0.0158	-0.0118
	(0.0123)	(0.0115)	(0.0120)
Age	0.0072***	0.0072***	0.0066***
	(0.0023)	(0.0021)	(0.0022)
Married	0.0095	0.0059	0.0053
	(0.0100)	(0.0111)	(0.0107)
Has children ages 0-5	0.1524***	0.1766***	0.1683***
	(0.0531)	(0.0568)	(0.0529)
Has children age 6-17	0.0287***	0.0341***	0.0308***
	(0.0092)	(0.0097)	(0.0093)
Spouse's income/1000	0.00026**	0.00024**	0.00018*
	(0.00011)	(0.00009)	(0.00010)
Urban	0.0014	0.0030	0.0036
	(0.0091)	(0.0088)	(0.0085)
AFQT	-0.0018	-0.0012	0.0003
	(0.0069)	(0.0060)	(0.0064)
Locus of control	-0.0048	-0.0050	-0.0049
	(0.0048)	(0.0034)	(0.0048)
Self-esteem	0.0041	0.0046	0.0047
	(0.0048)	(0.0043)	(0.0047)
South	-0.0102		
	(0.0083)		
Regional unemployment rate	0.0112**		
	(0.0057)		
Number of observations	33,648	33,614	33,648

Standard errors adjusted for serial correlation are in parentheses.

The reference category is 12 years of schooling.

All equations also include individual time means for marital status and spouse's income.

In column (1), the unemployment rate and south indicator are included in both employment and self-employment equations. In column (2), number of siblings in 1979 is used instead of the same gender indicator in the young children equation. In column (3), neither the number of siblings nor the same gender indicator is included in the young children equation.

Table 7: Robustness checks: Estimated average partial effects on the probability of self-employment when an alternative definition of d_{it} is used.

	1 Eq. Model	2 Eq. Model with s	2 Eq. Model with d	3 Eq. Model
	y Eq.	y Eq.	y Eq.	y Eq.
	(1)	(2)	(3)	(4)
Less than 12 years of schooling	0.0040	0.0116	0.0040	0.0111
	(0.0116)	(0.0183)	(0.0117)	(0.0168)
13 to 15 years of schooling	-0.0046	-0.0066	-0.0052	-0.0073
	(0.0092)	(0.0110)	(0.0093)	(0.0111)
More than 15 years of schooling	-0.0131	-0.0169	-0.0134	-0.0173
	(0.0095)	(0.0115)	(0.0096)	(0.0114)
Age	0.0046***	0.0055**	0.0049***	0.0059***
	(0.0016)	(0.0022)	(0.0016)	(0.0022)
Married	0.0221***	0.0244***	0.0185***	0.0202***
	(0.0060)	(0.0073)	(0.0067)	(0.0076)
Has children ages < 1	0.0188***	0.0265**	0.0927	0.1056
	(0.0072)	(0.0132)	(0.0636)	(0.0689)
Has children ages 1-5	0.0474***	0.0616***	0.0471***	0.0614***
	(0.0057)	(0.0199)	(0.0057)	(0.0171)
Has children age 6-17	0.0221***	0.0276***	0.0222***	0.0277***
	(0.0063)	(0.0093)	(0.0064)	(0.0089)
Spouse's income/1000	0.00016***	0.00024**	0.00016***	0.00024**
	(0.00006)	(0.00011)	(0.00006)	(0.00010)
Urban	0.0045	0.0039	0.0045	0.0037
	(0.0065)	(0.0076)	(0.0067)	(0.0078)
AFQT	0.0017	0.0002	0.0016	0.0001
	(0.0049)	(0.0062)	(0.0049)	(0.0061)
Locus of control	-0.0036	-0.0042	-0.0037	-0.0043
	(0.0037)	(0.0044)	(0.0037)	(0.0044)
Self-esteem	0.0043	0.0043	0.0042	0.0043
	(0.0037)	(0.0043)	(0.0038)	(0.0044)
Number of observations	28,446	33,648	28,446	33,648

Standard errors adjusted for serial correlation are in parentheses.

Indicator d_{it} is equal to one if the woman has a newborn child in year t, and is zero otherwise.

The reference category is 12 years of schooling.

All equations also include individual time means for marital status and spouse's income.