Revisiting the Composition of the Female Workforce - A Heckman Selection Model with Endogeneity

Jörg Schwiebert*

Abstract

In this paper, we revisit the Mulligan and Rubinstein (2008: Selection, investment and women's relative wages over time. *The Quarterly Journal of Economics*, 123(3):1061-1110) analysis about the composition of the female workforce in the United States. Using a Heckman selection model, these authors found that the selection of women into the female workforce changed from negative to positive over time. However, the authors assumed the exogeneity of covariates, which is sometimes appropriate but not for a variable like education. We revisit the issue of the Mulligan and Rubinstein (2008) paper by developing and applying a Heckman selection model which also controls for the potential endogeneity of education. Applying this estimator to U.S. Census and American Community Survey data, we find that selection has become more positive over time (like in Mulligan and Rubinstein), but that selection has never been negative. We rather find an interesting puzzle concerning the correlation pattern of the unobservables in our model which requires further investigation.

Keywords: Sample selection model, endogenous covariates, gender wage gap, composition of the female workforce, female labor force participation.
JEL codes: C21, C24, C26, J21, J31.

*Leibniz University Hannover, Institute of Labor Economics, Königsworther Platz 1, 30167 Hannover, Germany, Tel.: 0511/762-5657, E-mail: schwiebert@aoek.uni-hannover.de. Note: This paper replaces Discussion Paper No. 483 entitled "A Full Information Maximum Likelihood Approach to Estimating the Sample Selection Model with Endogenous Covariates", 2011, Department of Economics and Business Administration, Leibniz University Hannover.

1 Introduction

In their 2008 *QJE* paper, Mulligan and Rubinstein (2008) found that the composition of the female workforce in the United States has changed over time. In particular, they provided evidence that in the late 1970s working women were negatively selected from the female population, while in the late 1990s there was a positive selection of women into the female workforce. Positive selection means that those women with the highest skills are in the labor force, while the opposite is true in case of negative selection. The authors argue that the fact that more highly skilled women attended the labor force is able to explain the narrowing of the male-female wage gap in the United States over this time period.

Mulligan and Rubinstein employed the Heckman sample selection model to draw their conclusions. However, they did not control for the potential endogeneity of education, a variable which is likely to be endogenous. Ignoring the potential endogeneity of education, though, may lead to inconsistent parameter estimates so that the results of Mulligan and Rubinstein might be invalid. In this paper, we set up a Heckman sample selection model which also accounts for the potential endogeneity of education and revisit the issues addressed in the Mulligan and Rubinstein paper. In particular, we analyze whether accounting for the endogeneity of education has an effect on the amount and direction of sample selectivity of working women.

To see why education may be endogenous in the Mulligan/Rubinstein example, note that there might be unobservable factors like ability which are likely to jointly affect a woman's wage (main equation), her probability of labor force participation (selection equation) and her educational attainment. Since these unobservable factors cannot be included as control variables in our econometric model, we have a typical situation of endogeneity as the education variable will be correlated with the error terms of the main and the selection equation.[1]

[1] Recall that the Heckman selecton model is based on two equations, the main equation (of interest) and a selection equation which governs the probability of being selected.

Before analyzing the econometric implications of the joint presence of sample selectivity and endogeneity, we begin with a description of the virtues and drawbacks of the Heckman selection model. The Heckman selection model, originated in 1979, is based on a joint normality assumption of the error terms in the main equation and the selection equation. This bivariate normality assumption has often been criticized and challenged since. The reason is that parameter estimates of the main equation are typically inconsistent if the bivariate normality assumption is violated. Beginning in the 1980s, several authors have, thus, provided semi- and nonparametric estimation strategies which impose minimal assumptions on the distribution of error terms and are, therefore, more likely to consistently estimate the parameters of interest in applications. Examples include Gallant and Nychka (1987), Powell (1987), Ahn and Powell (1993), Das et al. (2003) and Newey (2009).

Most of these estimation approaches focus on consistent estimation of the parameters in the main equation and selection equation, where the latter ones can be viewed as nuisance parameters which are needed to obtain consistent estimates of the main equation parameters. However, if one focuses on the direction of selection and, thus, on the sample selection mechanism itself, one needs an estimation strategy which also gives an estimate of a parameter which measures the correlation between main and selection equation. If this correlation parameter is positive, we talk about positive selection, while in the other case we call this negative selection. The paper of Mulligan and Rubinstein is an example why estimation of a correlation parameter is useful, as it shows that the composition of the female workforce has changed from negative to positive selection, which in turn provides explanation why the male-female wage gap may have narrowed.

In this paper, we expand the Heckman selection model to allow for endogeneity of covariates. As in other econometric models, not accounting for endogeneity if it is indeed present leads to inconsistent parameter estimates.

First, we present a fully efficient full information maximum likelihood framework; thereafter, we present a limited information maximum likelihood framework which can be easily implemented in standard econometrics software. The latter does not give correct estimates of the correlation parameter, but appropriate formulas to calculate it are also provided.

Econometric estimators which account for sample selectivity and the presence of endogenous covariates have been provided by Das, Newey and Vella (2003), Chib et al. (2009) and Wooldridge (2010). Das, Newey and Vella (2003) employ a nonparametric estimation framework which relies on few distributional assumptions, but does not provide an estimate of a correlation parameter or a similar measure of the interdependence between main and selection equation. Chib et al. (2009) advocate a Bayesian framework and set up a full information maximum likelihood procedure. Our approach is quite similar, but easier to implement, and it is not Bayesian. Wooldridge (2010) suggests to augment the main equation with an inverse Mills ratio term and then to estimate this equation by two stage least squares (2SLS), thereby eliminating the endogeneity bias[2]. However, Wooldridge's strategy only works if the endogenous explanatory variable appears only in the main equation but not in the selection equation.

Our proposed likelihood estimator is in the spirit of the estimators for the Tobit model with endogenous covariates as provided by Smith and Blundell (1986) and the probit model with endogenous covariates as provided by Rivers and Vuong (1988); see also Newey (1987). Despite the fact that these estimators rely on strong distributional assumptions (as our estimator, too), they are implemented in standard econometric software packages (such as STATA) and are still frequently used in applied work. Even if one questions the parametric assumptions underlying these estimators, they still may be used for exploratory data analysis followed by a more appropriate procedure afterwards.

As pointed out above, and in contrast to the Smith-Blundell and Rivers-Vuong estimators, our estimator cannot simply be "improved" by a semi-/nonparametric estimator since it has the unique virtue that it provides an estimate of the correlation between main and selection equation, which is an important ingredient for analyzing how the sample se-

[2]This approach has been put forward in Semykina and Wooldridge (2010).

lection mechanism works. We will illustrate this latter point by revisiting the composition of the female workforce in the U.S. in the spirit of Mulligan and Rubinstein (2008).

The remainder of the paper is organized as follows. In section 2, we set up an econometric model which allows for the simultaneous presence of sample selectivity and endogeneity. Section 3 presents the full information and limited information maximum likelihood estimation strategies and shows how the latter can be implemented in standard econometric software. In this section, we also provide a test which indicates whether endogeneity is indeed present. In section 4, we apply our estimator to the Mulligan and Rubinstein analysis. Section 5 concludes the paper.

2 Econometric Model

The primary purpose of this paper is to revisit the Mulligan and Rubinstein example, where we seek to analyze the effects of treating education as endogenous. However, we also want to present a rather general framework for incorporating endogeneity into the Heckman selection model. The reason is that endogeneity may occur in three respects. First, endogeneity may only appear in the main but not in the selection equation; second, endogeneity may appear only in the selection but not in the main equation; and third, endogeneity may appear in both equations. Thus, we set up a relatively general model to cover all these cases. The model is given by

$$y_i^* = X_{1i}\beta_1 + X_{2i}\beta_2 + C_i\beta_3 + u_i \equiv X_i\beta + u_i \tag{1}$$

$$z_i^* = W_{1i}\gamma_1 + W_{2i}\gamma_2 + C_i\gamma_3 + Q_i\gamma_4 + v_i \equiv W_i\gamma + v_i \tag{2}$$

$$X_{2i} = [X_{1i}, W_{1i}]\Delta_1 + Z_{1i}\Delta_2 + \varepsilon_{1i} \equiv \tilde{Z}_{1i}\Delta + \varepsilon_{1i} \tag{3}$$

$$W_{2i} = [X_{1i}, W_{1i}]\Lambda_1 + Z_{2i}\Lambda_2 + \varepsilon_{2i} \equiv \tilde{Z}_{2i}\Lambda + \varepsilon_{2i} \tag{4}$$

$$C_i = [X_{1i}, W_{1i}]\Upsilon_1 + Z_{3i}\Upsilon_2 + \epsilon_{3i} \equiv \tilde{Z}_{3i}\Upsilon + \varepsilon_{3i} \tag{5}$$

$$z_i = 1(z_i^* > 0) \tag{6}$$

$$y_i = y_i^*\mathbf{1}(z_i = 1), \tag{7}$$

where $i = 1, \ldots, n$ indexes individuals. The first equation is the main equation, where the latent dependent variable y^* is related to a $(1 \times K_1)$-vector of exogenous explanatory variables, X_1, to a $(1 \times K_2)$-vector of endogenous explanatory variables only included in the main equation but not in the selection equation, X_2, and to a $(1 \times P)$-vector of endogenous explanatory variables included in the main and the selection equation, C. The second equation is the selection equation, where the latent variable z^* is related to a $(1 \times L_1)$-vector of exogenous explanatory variables, W_1, to a $(1 \times L_2)$-vector of endogenous explanatory variables, W_2 only included in the selection equation but not in the primary equation, to C and to Q. Q is an exogenous variable (it could also be a vector) which appears only in the selection equation. This is a well-known exclusion restriction serving to identify the parameters of the main equation. In equations (2.3) to (2.5) it is assumed that the endogenous explanatory variables can be explained by a $(1 \times M_1)$-vector, a $(1 \times M_2)$-vector and a $(1 \times M_3)$-vector of instrumental variables, Z_1, Z_2 and Z_3, respectively. Equation (2.6) expresses that only the sign of z^* is observable. Finally, equation (2.7) comprises the selection mechanism, i.e. the latent variable y^* is only observed if the selection indicator z is equal to one. Equations (2.1), (2.2), (2.6), and (2.7) build up the framework of the sample selection model without endogeneity as presented in many textbooks (e.g., Davidson and MacKinnon, 1993, pp. 542-543). The additional feature in equations (2.3) to (2.5) is that some of the covariates (X_2, W_2 and C) in the primary and the selection equation are endogenous, i.e. correlated with the error terms u and v. We assume that for each of these endogenous variables there exist instrumental variables Z_1, Z_2 and Z_3 which are not correlated with any error term in the model.

To complete the model, it is assumed that the vector of error terms $(u_i, v_i, \epsilon'_{1i}, \epsilon'_{2i}, \epsilon'_{3i})'$ is distributed according to

$$\begin{pmatrix} u_i \\ v_i \\ \varepsilon'_{1i} \\ \varepsilon'_{2i} \\ \varepsilon'_{3i} \end{pmatrix} \sim \text{NID} \left(0, \begin{bmatrix} \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} & \Omega' \\ \Omega_{(J\times 2)} & \Sigma_{(J\times J)} \end{bmatrix} \right), \tag{8}$$

where NID denotes "normally and independently distributed", $J \equiv K_2 + L_2 + P$, and the distribution should be interpreted as conditional on all exogenous variables (the conditioning has been omitted for the ease of notation). The covariance matrix of the error terms consists of four parts. The upper left part is the covariance matrix attributed to the error terms of the primary and selection equation, respectively, where σ_u^2 and σ_v^2 denote the variances of u and v, and ρ denotes the correlation coefficient. If there was no concern about endogeneity, inference would be based solely on this part of the covariance matrix, as it is common in the standard sample selection model. However, the (potential) presence of endogeneity is indicated by the $(J \times 2)$-matrix Ω, which captures the influence of unobserved factors which jointly affect the dependent variables in equation (2.1) and (2.2) and the endogenous explanatory variables. Note that endogeneity is absent if and only if Ω is equal to the null matrix. Finally, the error terms attributed to the endogenous explanatory variables have covariance matrix Σ whose dimension is $(J \times J)$.

Note that it is assumed that the distribution of the endogenous covariates can be reasonably approximated by a normal distribution, which favors continuous regressors and excludes binary regressors.

3 Estimation and Testing for Exogeneity

First, we lay out a full information maximum likelihood procedure in which all parameters of the model (1)-(6) are estimated simultaneously. First, note that the conditional distribution of $(u_i, v_i)'$ given $(\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i})$ is given by

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \Bigg| \varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i} \sim \text{NID}\left(\Omega'\Sigma^{-1} \begin{bmatrix} \varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i} \end{bmatrix}', B \right) \tag{9}$$

where

$$B \equiv \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} - \Omega'\Sigma^{-1}\Omega. \tag{10}$$

Define

$$\Psi \equiv \begin{pmatrix} \underset{(1\times K_2)}{\psi_{11}} & \underset{(1\times L_2)}{\psi_{12}} & \underset{(1\times P)}{\psi_{13}} \\ \underset{(1\times K_2)}{\psi_{21}} & \underset{(1\times L_2)}{\psi_{22}} & \underset{(1\times P)}{\psi_{23}} \end{pmatrix}_{(2\times J)} \equiv \Omega'\Sigma^{-1} \tag{11}$$

$$\Gamma \equiv \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix} \equiv \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix} - \Omega'\Sigma^{-1}\Omega, \tag{12}$$

where the lower right element of Γ has been set equal to unity due to normalization. Therefore, equation (3.1) can be recast as

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \Bigg| \varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i} \sim \text{NID}\left(\begin{bmatrix} \psi_{11}\varepsilon_{1i}' + \psi_{12}\varepsilon_{2i}' + \psi_{13}\varepsilon_{3i}' \\ \psi_{21}\varepsilon_{1i}' + \psi_{22}\varepsilon_{2i}' + \psi_{23}\varepsilon_{3i}' \end{bmatrix}, \begin{pmatrix} \tilde{\sigma}^2 & \tilde{\rho}\tilde{\sigma} \\ \tilde{\rho}\tilde{\sigma} & 1 \end{pmatrix} \right), \tag{13}$$

which resembles the (unconditional) joint error distribution of the sample selection model without endogeneity (except for the non-zero means).[3]

Then, the likelihood function can be written as the product of a conditional distribution which resembles the (unconditional) likelihood function of the sample selection model without endogeneity and the joint distribution of the error terms $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. Thus, the log-likelihood function is given by

$$\begin{aligned} l(\theta) = &\sum_{z_i=0} \log\{\Phi(-W_i\gamma - \psi_{21}\varepsilon'_{1i} - \psi_{22}\varepsilon'_{2i} - \psi_{23}\varepsilon'_{3i})\} \\ &+ \sum_{z_i=1} \log\{\tilde{\sigma}^{-1}\phi(\tilde{\sigma}^{-1}(y_i - X_i\beta - \psi_{11}\varepsilon'_{1i} - \psi_{12}\varepsilon'_{2i} - \psi_{13}\varepsilon'_{3i}))\} \\ &+ \sum_{z_i=1} \log\{\Phi((1-\tilde{\rho}^2)^{-1/2}[W_i\gamma + \psi_{21}\varepsilon'_{1i} + \psi_{22}\varepsilon'_{2i} + \psi_{23}\varepsilon'_{3i} \\ &\qquad + \tilde{\rho}\tilde{\sigma}^{-1}(y_i - X_i\beta - \psi_{11}\varepsilon'_{1i} - \psi_{12}\varepsilon'_{2i} - \psi_{13}\varepsilon'_{3i})])\} \\ &- \frac{n}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n} \begin{bmatrix} \varepsilon_{1i} & \varepsilon_{2i} & \varepsilon_{3i} \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \varepsilon_{1i} & \varepsilon_{2i} & \varepsilon_{3i} \end{bmatrix}', \end{aligned} \tag{14}$$

where $\theta \equiv (\beta', \gamma', \tilde{\rho}, \tilde{\sigma}, \text{vec}(\Psi)', \text{vech}(\Sigma)', \text{vec}(\Delta)', \text{vec}(\Lambda)', \text{vec}(\Upsilon)')'$,

$$\varepsilon_{1i} = X_{2i} - \tilde{Z}_{1i}\Delta \tag{15}$$

$$\varepsilon_{2i} = W_{2i} - \tilde{Z}_{2i}\Lambda \tag{16}$$

$$\varepsilon_{3i} = C_i - \tilde{Z}_{3i}\Upsilon, \tag{17}$$

$\Phi(\cdot)$ denotes the standard normal cumulative distribution function and $\phi(\cdot)$ the standard normal probability density function.

The FIML estimator of the sample selection model with endogenous covariates is thus given by

$$\hat{\theta} = \arg\max_{\theta}\ l(\theta). \tag{18}$$

[3]The approach undertaken here to accommodate the endogeneity problem is known as a "control function approach" in the literature (see, e.g., Wooldridge, 2010, pp. 126-29).

The FIML estimator actually does not provide estimates of the "structural" variance-covariance parameters, i.e., those parameters in the unconditional distribution of the error terms. Especially the correlation parameter between main and selection equation is of interest, as pointed out in the introduction. However, these structural parameters can be deduced from the FIML estimates by noting that

$$\hat{\Pi} = \hat{\Gamma} + \hat{\Psi}\hat{\Sigma}\hat{\Psi}' \tag{19}$$

$$\hat{\Omega} = \hat{\Sigma}\hat{\Psi}' \tag{20}$$

$$\hat{\rho} = \frac{\hat{g}}{\hat{\sigma}_u \hat{\sigma}_v}, \tag{21}$$

where $\Pi \equiv \begin{pmatrix} \sigma_u^2 & g \\ g & \sigma_v^2 \end{pmatrix}$ and $g \equiv \rho\sigma_u\sigma_v$. In the appendix it is shown how standard errors for these structural estimates can be derived by means of the delta method.[4]

The FIML estimator is fully efficient. However, if the number of observations is large and/or the number of covariates is large, estimation may be quite time consuming. As an alternative, one may consider choosing a limited maximum likelihood (LIML) approach. We propose the following procedure:

1) Estimate the reduced form equations (3)-(5) by OLS and obtain the residuals $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$.

[4] We also provide in the appendix a small Monte Carlo simulation study which analyzes the finite sample performance of the FIML estimator and compares its estimates to the (biased) estimates based on the ordinary Heckman selection model which does not control for endogeneity. Moreover, we provide an application of our estimator to the well-known Mroz (1987) labor supply data set in order to compare our results with those of Wooldridge (2010), who did the same using his estimator.

2) Insert these estimated values into the following log-likelihood function

$$
\begin{aligned}
l(\tilde{\theta}) = & \sum_{z_i=0} \log\{\Phi(-W_i\gamma - \psi_{21}\hat{\varepsilon}'_{1i} - \psi_{22}\hat{\varepsilon}'_{2i} - \psi_{23}\hat{\varepsilon}'_{3i})\} \\
& + \sum_{z_i=1} \log\{\tilde{\sigma}^{-1}\phi(\tilde{\sigma}^{-1}(y_i - X_i\beta - \psi_{11}\hat{\varepsilon}'_{1i} - \psi_{12}\hat{\varepsilon}'_{2i} - \psi_{13}\hat{\varepsilon}'_{3i}))\} \\
& + \sum_{z_i=1} \log\{\Phi((1-\tilde{\rho}^2)^{-1/2}[W_i\gamma + \psi_{21}\hat{\varepsilon}'_{1i} + \psi_{22}\hat{\varepsilon}'_{2i} + \psi_{23}\hat{\varepsilon}'_{3i} \\
& \qquad + \tilde{\rho}\tilde{\sigma}^{-1}(y_i - X_i\beta - \psi_{11}\hat{\varepsilon}'_{1i} - \psi_{12}\hat{\varepsilon}'_{2i} - \psi_{13}\hat{\varepsilon}'_{3i})])\},
\end{aligned} \tag{22}
$$

which is then maximized over $\tilde{\theta} \equiv (\beta', \gamma', \tilde{\rho}, \tilde{\sigma}, \mathrm{vec}(\Psi)')$.

Observe that the log-likelihood function is the same as for the Heckman selection model without endogeneity, with the difference that we have the additional covariates $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$. Thus, our model can be estimated using any econometrics software which supports maximum likelihood estimation of the Heckman selection model. One must simply add to the set of covariates the estimated residuals $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$.

Of course, using estimated residuals as covariates instead of the true error terms requires an adjustment of the (asymptotic) standard errors. To get appropriate standard errors, one can either

a) use a correction formula which gives that $\sqrt{n}(\hat{\tilde{\theta}} - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, C)$, where C is the corrected asymptotic covariance matrix which accounts for the estimation error in $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$. The exact expression for C is provided in the appendix;

b) combine the first order conditions from maximizing the limited information log-likelihood function with the normal equations for estimating the reduced form equations for the endogenous explanatory variables and estimate the parameters jointly in a generalized method of moments (GMM) framework;

c) use the bootstrap.

We now present a simple test which indicates whether endogeneity is indeed a problem in a particular application. The absence of endogeneity means that the matrix Ω is equal to the null matrix. But this implies that Ψ is equal to the null matrix as well. Hence, we can test for the absence of endogeneity by performing a simple test of joint significance of the parameters associated with the additional "covariates" $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$ and $\hat{\varepsilon}_3$. If we cannot reject the joint hypothesis that these parameters are equal to zero, then we may conclude that endogeneity is indeed absent and estimates from an ordinary Heckman selection model would be consistent.

A test of the hypothesis that $\Psi = 0$ is a standard task in maximum likelihood estimation. We propose a Wald test. In that case, the test statistic will be given by

$$W_\Psi = \text{vec}(\hat{\Psi})'(\text{Asy.Cov}[\text{vec}(\hat{\Psi})])^{-1}\text{vec}(\hat{\Psi}) \sim \chi^2(2J), \tag{23}$$

where Asy.Cov[vec($\hat{\Psi}$)] denotes the asymptotic covariance matrix of vec($\hat{\Psi}$). In case of the FIML estimator and provided that suitable regularity conditions hold (for instance, cf. Amemiya, 1985, pp. 120-127), this asymptotic covariance can be obtained by using the fact that

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, -\mathcal{H}^{-1}), \tag{24}$$

where $\mathcal{H} = n^{-1}\text{E}\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta'}\right)$.

4 Empirical Analysis

In this section, we apply our LIML estimator to the Mulligan and Rubinstein (2008) example. Mulligan and Rubinstein (2008) used CPS data to fit wage regressions for men and women. The estimates for women were conducted using the two-step Heckman selection estimator which controls for sample selectivity. Based on their estimation results, the authors concluded that selection of women into the labor force had turned over time from negative to positive, a fact which might explain the narrowing of the male-female wage gap over the time period under consideration.

We do basically the same as Mulligan and Rubinstein (2008) did. More specifically, we estimate wage regressions for women and analyze the composition of the female workforce while controlling for sample selectivity *and* endogeneity. Instead, however, of using CPS data, we employ the public use files of the 1980 US Census and the 2005-2010 American Community Survey (ACS). The reason is that we need plausible instrumental variables for education. These should be randomly assigned, affect the wage only through the effect on education ("exclusion") and should have a statistically significant relation to education ("first stage"). Instrumental variables satisfying these conditions are hard to find.

To resolve this issue, we exploit the idea underlying the Angrist and Krueger (1991) paper. Angrist and Krueger (1991) used the quarter of birth (and various interactions) as an instrumental variable for education. The idea is that children in the United States attend school in the year they turn six, where December the 31st is the cutoff date. Thus, a child who turns six late in the year attends school at the age of five, whereas a child who turns six early in the year attends school at the age of six. Since the legal high school drop out age in the United States is 16 years of age, Angrist and Krueger (1991) argue that children born late in the year attend school at an earlier age and, thus, stay longer in school.

Unfortunately, a quarter of birth variable is not included in the CPS files for the entire time period of interest. We thus employ the 1980 US Census and the American Community Survey data from 2005 to 2010, since these data sets do contain the quarter of birth variable.[5]

Mulligan and Rubinstein (2008) perform regressions for the time periods 1975-1979 and 1995-1999, which means they pooled these years and estimated wage regressions for men and women. We will estimate wage regressions only for women since we are interested in the composition of the female workforce. Moreover, we take the 1980 Census as a substitute for the period 1975-1979 and the (pooled) ACS files from 2005-2010 as a substitute for 1995-1999. Despite the fact that the ACS files cover a period ten years

[5]We obtained our data files from the IPUMS-USA database (Ruggles et al., 2010).

after, we conjecture to find the same basic pattern, i.e., that the selection of women into the workforce has become more positive over time.

As in Mulligan and Rubinstein (2008), our sample consists of white non-Hispanic adults between 25 and 54 years of age not living in group quarters. Moreover, our sample includes women only. As the working population we only consider full time full year (FTFY) workers, i.e., workers who worked at least 36 hours per week and 50 hours in the last year. Only for these women we calculated an hourly wage given by their annual income divided by (52 times the usual hours of work).[6] The remaining women add to the non-selected population, which thus comprises women who do not work at all and women who did not work full time full year. Hence, the selection parameter (i.e., the correlation between main and selection equation) that we will compute refers to the selection of women into the full time full year workforce. Put differently, the subject of our analysis is the composition of the full time full year workforce in 1980 and from 2005 to 2010.

In order to prevent our estimation results from the impact of outliers, we excluded incomes below the fifth percentile and above the 95th percentile in each time period. Moreover, observations for which incomes have been imputed by a "hot deck" procedure were eliminated as well. We also excluded unemployed people as we cannot say whether these (potentially) belong to the FTFY people or to the remaining population. Furthermore, we eliminated self-employed workers.

Due to Mulligan and Rubinstein (2008), we specify the following empirical model. The main equation has the natural logarithm of the hourly wage as its dependent variable, so that the estimated coefficients of the explanatory variables can be interpreted as the percentage change in the wage rate due to a one-unit increase in an explanatory variable (in case of continuous variables). Covariates in the main equation include educational attainment (*educ*), age (*age*), age squared (*age2*), dummies for the census region (*northeast*, *midwest*, *south*; *west* is the baseline) and dummies for the marital status (*widowed*,

[6] Of course, these wage calculations have been adjusted for the inflation rate. We multiplied each hourly wage with the cpi99 variable included in the public use samples, which expresses all nominal values in 1999 US dollars.

divorced, *separated*, *never married*; *married* is the baseline). The selection equation includes the same variables as the main equation and the number of children younger than five years of age (*nchlt5*). The latter variable is exclusive in the selection equation and serves for the sake of identification.

Since education is potentially endogenous, we follow the LIML approach outlined in section 3 and estimate in the first stage a reduced form equation for education. Explanatory variables are the exogenous variables from the main equation and our quarter of birth dummies (where the first quarter is the baseline) as instrumental variables.

Table 1 provides descriptive statistics of the variables. We see that the people from the pooled ACS samples have a higher wage on average, have more educational attainment, are older and have a lower probability of being married. The quarter of birth dummies are almost evenly distributed over the population, as one would have expected. In the 1980 Census file we have 1,032,668 observations, while we have 1,528,735 observations in the pooled ACS files. In the 1980 Census file 40.04 percent of these people worked, while in the ACS files 62.06 percent worked. We used these observations unweighted in our regressions; results were very similar for the weighted and unweighted samples.

We begin our empirical analysis with the reduced form estimates for education. From table 2 we can see that for both samples the quarter of birth dummies have a significant impact on the education variable, thus fulfilling one basic requirement to be valid instrumental variables. In the 1980 sample, the coefficients on these dummies possess the expected signs, since the coefficient values imply that the educational attainment of people born late in the year is higher. For the 2005-2010 sample, however, we have the puzzling finding that people born in the second quarter of the year have the highest educational attainment. This can also be seen in the regression-unadjusted means of education by year and quarter of birth, as provided in table 3. The estimated coefficients for the quarter of birth dummies for the 2005-2010 sample are still significantly different from zero, though, and we maintain our assumption that the quarter of birth dummies are exogenous and, thus, valid instrumental variables.

From these first stage estimates, we obtain the estimated residuals (*eps*) and insert them as additional covariates into a maximum likelihood estimation procedure of the Heckman selection model. By doing this, we not only control for sample selectivity but for endogeneity as well. But before we proceed so far, we estimate a Heckman selection model which does not account for endogeneity. This is the approach chosen by Mulligan and Rubinstein (2008), and we expect to find (qualitatively) similar results using our data sets.

The results for the ordinary Heckman model without controlling for endogeneity can be found in table 4. The most important result from table 4 for our study is the estimated value of the correlation coefficient between main and selection equation. We see that the coefficient is negative for the 1980 sample (albeit not significantly different from zero) and positive for the 2005-2010 sample. Hence, these estimates reflect the main findings in Mulligan and Rubinstein (2008) that the selection has become more positive over time and that it has changed from negative to positive.

When controlling for the potential endogeneity of education, we obtain the results shown in table 5. The first result to highlight is that the returns to education are nearly doubled when endogeneity is taken into account. But more important, the correlation coefficient has increased for both samples. We still see an increase of the correlation coefficient over time, so one of the Mulligan/Rubinstein (2008) findings holds true when we control for endogeneity. However, we also see that the correlation is positive in 1980. Thus, these results do not support the hypothesis that selection of women into the workforce was negative in 1980. On the contrary, these results provide evidence that the unobserved factors which lead to an increased probability of labor force participation also have a positive impact on the wage.

Moreover, we find that the estimated coefficients of *eps* are significantly different from zero in both main and selection equation, which indicates that an endogeneity bias is present. We also find that these coefficients are negative. Since the coefficients measure the correlation between unobserved factors governing educational attainment and unobserved factors in main and selection equation, we are faced with an interesting puzzle. There seem to be latent factors which increase the probability of labor force participation and increase wages, but which decrease educational attainment. Thus the plain ability-like story from the introduction cannot be true, since if ability was dominant among the unobserved factors, then the unobserved factors should be positively correlated. It remains an interesting question for future research why the correlation structure of unobservables has the pattern found in our estimation results.

5 Conclusions

In this paper, we have revisited the Mulligan and Rubinstein (2008) analysis of the composition of the female workforce. Using a different data set as in the original analysis, we also find that the selection of women into the female workforce has become more positive over time, which remains true whether or not we control for the endogeneity of education. However, when taking endogeneity of education into account, we do not have a switch from negative to positive selection over time any more. On the contrary, we find an interesting relationship between the unobservables of our econometric model. While these unobserved factors are positively correlated between main and selection equation, they are negatively correlated with the unobserved factors governing educational attainment.

To reach our conclusions we employed a Heckman sample selection model with endogenous covariates. We provided a rather general model which encompasses various scenarios of endogeneity, including endogeneity only in the main equation, only in the selection equation or in both. Although our estimator relies on distributional assumptions which may not be satisfied in applications, the estimator nevertheless serves as a starting point for a deeper (semiparametric) analysis. The main virtue of our estimator is that it is relatively simple to compute. In fact, any econometrics software which is capable of performing maximum likelihood estimation of the Heckman sample selection model can be used.

Despite its frequent use in applied econometrics, in most cases authors assume exogeneity of covariates when employing the Heckman selection model. The empirical analysis of the preceding section, however, has shown for instance that the returns to education are nearly doubled when accounting for endogeneity. This underlines the necessity of controlling for the joint presence of sample selectivity and endogeneity of covariates if one seeks to get consistent parameter estimates. This is especially important for analyzing group differences, e.g., wage differences between men and women. Decomposition methods used to analyze these differences, such as the well-known Blinder (1973)-Oaxaca (1973) decomposition, are only valid if the model parameters have been estimated consistently.

References

Ahn, H. and Powell, J. L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics*, 58(1-2):3–29.

Amemiya, T. (1985). *Advanced Econometrics*. Basil Blackwell, Oxford.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):pp. 979–1014.

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(4):pp. 436–455.

Card, D. (1999). The causal effect of education on earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 3 of *Handbook of Labor Economics*, chapter 30, pages 1801–1863. Elsevier.

Chib, S., Greenberg, E., and Jeliazkov, I. (2009). Estimation of semiparametric models in the presence of endogeneity and sample selection. *Journal of Computational and Graphical Statistics*, 18(2):321–348.

Das, M., Newey, W. K., and Vella, F. (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies*, 70(1):33–58.

Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press, New York, NY.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):pp. 363–390.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–61.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55(4):765–99.

Mulligan, C. B. and Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time. *The Quarterly Journal of Economics*, 123(3):1061–1110.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36(3):231–250.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *Econometrics Journal*, 12:S217–S229.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Powell, J. L. (1987). Semiparametric estimation of bivariate limited dependent variable models. Manuscript, University of California, Berkeley.

Rivers, D. and Vuong, Q. H. (1988). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, 39(3):347–366.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., and Sobek, M. (2010). Integrated public use microdata series: Version 5.0 [machine-readable database]. Minneapolis: University of Minnesota.

Semykina, A. and Wooldridge, J. M. (2010). Estimating panel data models in the presence of endogeneity and selection. *Journal of Econometrics*, 157(2):375–380.

Smith, R. J. and Blundell, R. W. (1986). An exogeneity test for a simultaneous equation tobit model with an application to labor supply. *Econometrica*, 54(3):679–85.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, Cambridge, MA, 2nd edition.

Appendix A

In this appendix, we show how the asymptotic covariance matrix of the LIML estimator must be corrected in order to account for the estimation of the regressors ε_1, ε_2 and ε_3. First, let $\alpha \equiv (\text{vec}(\Delta)', \text{vec}(\Lambda)', \text{vec}(\Upsilon)')'$ and $\tilde{l}(\tilde{\theta}, \hat{\alpha}) = \sum_{i=1}^{n} l_i(\tilde{\theta}, \hat{\alpha})$ be the limited information log-likelihood function. Provided there exists an interior solution, we can write the first order condition from maximizing this likelihood function as

$$\sum_{i=1}^{n} \frac{\partial l_i(\hat{\tilde{\theta}}, \hat{\alpha})}{\partial \tilde{\theta}} = 0. \tag{25}$$

An asymptotic first order expansion about $\hat{\tilde{\theta}} = \tilde{\theta}$ gives after rearranging and pre-multiplication with $\sqrt{n}$

$$\sqrt{n}(\hat{\tilde{\theta}} - \tilde{\theta}) = \left(-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}^2} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial l_i(\tilde{\theta}, \hat{\alpha})}{\partial \tilde{\theta}} + o_p(1). \tag{26}$$

Expanding the gradient about $\hat{\alpha} = \alpha$ yields

$$\begin{aligned} \sqrt{n}(\hat{\tilde{\theta}} - \tilde{\theta}) = & \left(-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}^2} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}} \\ & + \left(-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}^2} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta} \partial \hat{\alpha}} \right) \sqrt{n}(\hat{\alpha} - \alpha) + o_p(1). \end{aligned} \tag{27}$$

If

$$-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}^2} \xrightarrow{p} H \text{ pos. def.} \tag{28}$$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta}} \xrightarrow{d} \mathcal{N}(0, M) \tag{29}$$

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 l_i(\tilde{\theta}, \alpha)}{\partial \tilde{\theta} \partial \hat{\alpha}} \xrightarrow{p} J \tag{30}$$

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, V), \tag{31}$$

then

$$\sqrt{n}(\hat{\tilde{\theta}} - \tilde{\theta}) \xrightarrow{d} \mathcal{N}(0, C), \tag{32}$$

where $C = H^{-1}(M + JVJ')H^{-1}$. This follows because the covariance between $\frac{\partial l_i(\tilde{\theta},\alpha)}{\partial \tilde{\theta}}$ and $(\hat{\alpha} - \alpha)$ is zero, as shown by Smith and Blundell (1986).

Note that implementation of the LIML estimator using an econometrics software yields an asymptotic covariance of $H^{-1}MH^{-1}$, as the software does not know that some regressors have been estimated. Hence, one must add to this expression a correction term of $H^{-1}(JVJ')H^{-1}$ in order to obtain the correct asymptotic covariance.

Appendix B

In this appendix, we derive formulas for the (asymptotic) variances of the estimates of the structural variance-covariance parameters (based on the FIML estimates). We assume, however, that FIML estimation does not yield estimates of $\tilde{\rho}$, $\tilde{\sigma}$ and Σ, but rather of $\operatorname{atanh}(\tilde{\rho})$, $\ln(\tilde{\sigma})$ and S such that $\Sigma = SS'$. The reason for not directly estimating these parameters is that we have to make sure that $\hat{\tilde{\rho}} \in (-1, 1)$, $\hat{\tilde{\sigma}} > 0$ and $\hat{\Sigma}$ be positive definite. Our reparameterization guarantees that these conditions are fulfilled.

(i) The Asymptotic Distribution of $\hat{\Omega}$

ML estimation yields estimates of[7]

$$s \equiv \begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \end{bmatrix} = \text{vech}(S) \quad \text{and} \quad \text{vec}(\Psi') = \begin{bmatrix} \psi'_{11} \\ \psi'_{12} \\ \psi'_{21} \\ \psi'_{22} \end{bmatrix}_{(2J\times 1)} . \tag{33}$$

Let

$$q \equiv (s', \text{vec}(\Psi')')'. \tag{34}$$

Since $\Omega = \Sigma\Psi'$ is a function of q, the asymptotic distribution of $\text{vec}(\hat{\Omega})$ can be obtained by means of the Delta method. If

$$\sqrt{n}(\hat{q} - q) \xrightarrow{d} N(0, M), \tag{35}$$

then

$$\sqrt{n}(\text{vec}(\hat{\Omega}) - \text{vec}(\Omega)) \xrightarrow{d} N\left(0, CMC'\right), \tag{36}$$

where

$$\begin{aligned} C &= \frac{\partial \text{vec}(\Omega)}{\partial q'} && (37) \\ &= \frac{\partial \text{vec}(SS'\Psi')}{\partial q'} && (38) \\ &= (\Psi \otimes I_J)\frac{\partial \text{vec}(SS')}{\partial q'} + (I_2 \otimes SS')\frac{\partial \text{vec}(\Psi')}{\partial q'} && (39) \\ &= (\Psi \otimes I_J)\begin{bmatrix} \frac{\partial \text{vec}(SS')}{\partial s'} & 0 \end{bmatrix} + (I_2 \otimes \Sigma)\begin{bmatrix} 0 & \frac{\partial \text{vec}(\Psi')}{\partial \text{vec}(\Psi')'} \end{bmatrix} && (40) \\ &= \begin{bmatrix} (\Psi \otimes I_J)\frac{\partial \text{vec}(SS')}{\partial s'} & (I_2 \otimes \Sigma) \end{bmatrix}. && (41) \end{aligned}$$

[7]Note: The 2-by-2 case has been used here for the sake of illustration. The following analysis does not hinge on this case.

Furthermore,

$$\frac{\partial \text{vec}(SS')}{\partial s'} = \left\{(S \otimes I_J)\frac{\partial \text{vec}(S)}{\partial s'} + (I_J \otimes S)\frac{\partial \text{vec}(S')}{\partial s'}\right\} \tag{42}$$

$$= \left\{(S \otimes I_J)\frac{\partial \text{vec}(S)}{\partial s'} + (I_J \otimes S)K_J\frac{\partial \text{vec}(S)}{\partial s'}\right\} \tag{43}$$

$$= \{(S \otimes I_J)L'_J + (I_J \otimes S)K_J L'_J\} \tag{44}$$

$$= \{(S \otimes I_J) + (I_J \otimes S)K_J\} L'_J \tag{45}$$

$$= (I_{J^2} + K_J)(S \otimes I_J)L'_J, \tag{46}$$

with

$$L_J = \sum_{i \geq j} u_{ij}\text{vec}(E_{ij})' \tag{47}$$

$$K_J = \sum_{i=1}^{J}\sum_{j=1}^{J} E_{ij} \otimes E'_{ij}, \tag{48}$$

where u_{ij} denotes a unit vector of size $\frac{1}{2}J(J+1)$ whose $[(j-1)J+i-\frac{1}{2}j(j-1)]$-th element is unity $(1 \leq j \leq i \leq J)$, and E_{ij} is a $(J \times J)$ matrix with one at the (i,j)-th position and zeros elsewhere. Note that L_J and K_J do only depend on J.

Therefore,

$$C = \left[(\Psi \otimes I_J)(I_{J^2} + K_J)(S \otimes I_J)L'_J \quad (I_2 \otimes \Sigma)\right]. \tag{49}$$

(ii) The Asymptotic Distribution of $\hat{\Pi} = \begin{pmatrix} \hat{\sigma}_u^2 & \rho\hat{\sigma}_u\hat{\sigma}_v \\ \rho\hat{\sigma}_u\hat{\sigma}_v & \hat{\sigma}_v^2 \end{pmatrix}$

ML estimation yields estimates of

$$s \equiv \mathrm{vech}(S), \quad \mathrm{vec}(\Psi') = \begin{bmatrix} \psi'_{11} \\ \psi'_{12} \\ \psi'_{21} \\ \psi'_{22} \end{bmatrix}, \quad [\ln\tilde{\sigma}], \quad [\mathrm{atanh}(\tilde{\rho})]. \tag{50}$$

Let

$$q \equiv (s', \mathrm{vec}(\Psi')', [\ln\tilde{\sigma}], [\mathrm{atanh}(\tilde{\rho})])'. \tag{51}$$

Since

$$\Pi = \Gamma + \Psi\Sigma\Psi' \tag{52}$$

$$= \begin{bmatrix} (\exp\{[\ln\tilde{\sigma}]\})^2 & \tanh([\mathrm{atanh}(\tilde{\rho})])\exp\{[\ln\tilde{\sigma}]\} \\ \tanh([\mathrm{atanh}(\tilde{\rho})])\exp\{[\ln\tilde{\sigma}]\} & 1 \end{bmatrix} + \Psi\Sigma\Psi' \tag{53}$$

is a function of q, the asymptotic distribution of $\mathrm{vech}(\hat{\Pi})$ can be obtained by means of the delta method.

If

$$\sqrt{n}(\hat{q} - q) \xrightarrow{d} N(0, M), \tag{54}$$

then

$$\sqrt{n}(\mathrm{vech}(\hat{\Pi}) - \mathrm{vech}(\Pi)) \xrightarrow{d} N\left(0, CMC'\right), \tag{55}$$

where

$$
\begin{aligned}
C &= \frac{\partial \mathrm{vech}(\Pi)}{\partial q'} && (56)\\
&= L_{2J} \frac{\partial \mathrm{vec}(\Pi)}{\partial q'} && (57)\\
&= L_{2J} \left\{ \frac{\partial \mathrm{vec}(\Gamma)}{\partial q'} + \frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial q'} \right\}. && (58)
\end{aligned}
$$

Both components of the RHS have to be investigated in detail.

First,

$$
\frac{\partial \mathrm{vec}(\Gamma)}{\partial [\ln \tilde{\sigma}], [\mathrm{atanh}(\tilde{\rho})]} = \begin{bmatrix} 2(\exp\{[\ln \tilde{\sigma}]\})^2 & 0 \\ \tanh([\mathrm{atanh}(\tilde{\rho})]) \exp\{[\ln \tilde{\sigma}]\} & (1 - \tanh^2([\mathrm{atanh}(\tilde{\rho})])) \exp\{[\ln \tilde{\sigma}]\} \\ \tanh([\mathrm{atanh}(\tilde{\rho})]) \exp\{[\ln \tilde{\sigma}]\} & (1 - \tanh^2([\mathrm{atanh}(\tilde{\rho})])) \exp\{[\ln \tilde{\sigma}]\} \\ 0 & 0 \end{bmatrix} \tag{59}
$$

$$
\begin{aligned}
&\equiv A && (60)\\
\Rightarrow \frac{\partial \mathrm{vec}(\Gamma)}{\partial q'} &= \begin{bmatrix} 0 & A \end{bmatrix}. && (61)
\end{aligned}
$$

Next,

$$
\begin{aligned}
\frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} &= (\Psi\Sigma \otimes I_2) \frac{\partial \mathrm{vec}(\Psi)}{\partial (s', \mathrm{vec}(\Psi')')} + (I_2 \otimes \Psi) \frac{\partial \mathrm{vec}(\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} && (62)\\
&= (\Psi\Sigma \otimes I_2) K_{2J} \frac{\partial \mathrm{vec}(\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} + (I_2 \otimes \Psi) \frac{\partial \mathrm{vec}(\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} && (63)\\
&= (\Psi\Sigma \otimes I_2) K_{2J} \begin{bmatrix} 0 & I_{2J} \end{bmatrix} + (I_2 \otimes \Psi) \frac{\partial \mathrm{vec}(\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} && (64)\\
\Rightarrow \frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial q'} &= \begin{bmatrix} \frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} & 0 \end{bmatrix} && (65)
\end{aligned}
$$

Hence,

$$
C = \frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial q'} = \begin{bmatrix} \frac{\partial \mathrm{vec}(\Psi\Sigma\Psi')}{\partial (s', \mathrm{vec}(\Psi')')} & A \end{bmatrix}. \tag{66}
$$

(iii) The Asymptotic Distribution of $\hat{\rho}$

Given an estimate of

$$\Pi = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{pmatrix}, \tag{67}$$

let

$$g \equiv \rho\sigma_u\sigma_v \Rightarrow \rho = \frac{g}{\sigma_u\sigma_v} = \frac{g}{\sqrt{\sigma_u^2\sigma_v^2}} = g\left(\sigma_u^2\sigma_v^2\right)^{-\frac{1}{2}} \tag{68}$$

and

$$q \equiv \left(\sigma_u^2, g, \sigma_v^2\right)' = \text{vech}(\Pi). \tag{69}$$

Since ρ is a function of q, the asymptotic distribution of $\hat{\rho}$ can be obtained by means of the delta method.

If

$$\sqrt{n}(\hat{q} - q) \xrightarrow{d} \mathcal{N}(0, G) \tag{70}$$

then

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} \mathcal{N}(0, FGF') \tag{71}$$

with

$$F = \frac{\partial \rho}{\partial q'} = \left[-\frac{1}{2} g \left(\sigma_u^2 \sigma_v^2 \right)^{-\frac{3}{2}} \sigma_v^2, \left(\sigma_u^2 \sigma_v^2 \right)^{-\frac{1}{2}}, -\frac{1}{2} g \left(\sigma_u^2 \sigma_v^2 \right)^{-\frac{3}{2}} \sigma_u^2 \right]. \tag{72}$$

(iv) The Asymptotic Distribution of $\hat{\Sigma} = \hat{S}\hat{S}'$

ML estimation yields estimates of[8]

$$s \equiv \begin{bmatrix} s_{11} \\ s_{21} \\ s_{22} \end{bmatrix} = \text{vech}(S). \tag{73}$$

The asymptotic distribution is given by

$$\sqrt{n}(\hat{s} - s) \xrightarrow{d} N(0, M). \tag{74}$$

Since $\text{vech}(\Sigma) = \text{vech}(SS') = c(s)$ is a function of s, the asymptotic distribution of $\text{vech}(\Sigma)$ can be obtained by using the delta method, which gives

$$\sqrt{n}(\text{vech}(\hat{\Sigma}) - \text{vech}(\Sigma)) \xrightarrow{d} N(0, C(s)MC(s)'), \tag{75}$$

[8]Note: The 2-by-2 case has been used here for the sake of illustration. The following analysis does not hinge on this case.

where

$$
\begin{aligned}
C(s) &= \frac{\partial c(s)}{\partial s'} && (76)\\
&= \frac{\partial \text{vech}(SS')}{\partial s'} && (77)\\
&= L_J \frac{\partial \text{vec}(SS')}{\partial s'} && (78)\\
&= L_J \left\{ (S \otimes I_J) \frac{\partial \text{vec}(S)}{\partial s'} + (I_J \otimes S) \frac{\partial \text{vec}(S')}{\partial s'} \right\} && (79)\\
&= L_J \left\{ (S \otimes I_J) \frac{\partial \text{vec}(S)}{\partial s'} + (I_J \otimes S) K_J \frac{\partial \text{vec}(S)}{\partial s'} \right\} && (80)\\
&= L_J \left\{ (S \otimes I_J) L_J' + (I_J \otimes S) K_J L_J' \right\} && (81)\\
&= L_J \left\{ (S \otimes I_J) + (I_J \otimes S) K_J \right\} L_J' && (82)\\
&= L_J (I_{J^2} + K_J)(S \otimes I_J) L_J' && (83)
\end{aligned}
$$

and

$$
\begin{aligned}
L_J &= \sum_{i \geq j} u_{ij} \text{vec}(E_{ij})' && (84)\\
K_J &= \sum_{i=1}^{J} \sum_{j=1}^{J} E_{ij} \otimes E_{ij}', && (85)
\end{aligned}
$$

where u_{ij} denotes a unit vector of size $\frac{1}{2}J(J+1)$ whose $[(j-1)J+i-\frac{1}{2}j(j-1)]$-th element is unity ($1 \leq j \leq i \leq J$), and E_{ij} is a ($J \times J$) matrix with one at the (i,j)-th position and zeros elsewhere. Note that L_J and K_J do only depend on J.

Appendix C

In this appendix, we use Monte Carlo simulations in order to study the finite sample properties of our FIML estimator and in order to gauge the bias which occurs if one does not account for endogeneity. The results of these simulations are presented in table 6.

The first column of table 6 contains the specification. We distinguish between four benchmark cases. In the first case, endogeneity is only present in the primary equation.

In particular, it is assumed that

$$
\begin{array}{lllllll}
y_i^* & = .2 & +.4\,X_{1i} & +.9\,X_{2i} & & & +u_i \\
z_i^* & = 1 & & & +.7\,W_{1i} & & +v_i \\
X_{2i} & = .5 & +1.5\,X_{1i} & & -.2\,W_{1i} & +.7\,Z_{1i} & +\varepsilon_{1i}
\end{array}
$$

and

$$
\mathrm{Cov}[(u_i, v_i, \varepsilon_{1i})'] = \begin{pmatrix} 1 & & \\ .9 & 1 & \\ .5 & .4 & 2 \end{pmatrix}.
$$

Note that we have assumed a relatively high correlation between the primary and the selection equation. Hence, we focus our attention on situations where sample selection bias is indeed a problem.

In the second case, endogeneity is only present in the selection equation:

$$
\begin{array}{llllll}
y_i^* & = .2 & +.4\,X_{1i} & & & +u_i \\
z_i^* & = 1 & +.7\,X_{1i} & +.3\,W_{2i} & & +v_i \\
W_{2i} & = .5 & +1.5\,X_{1i} & & +.7\,Z_{2i} & +\varepsilon_{2i}
\end{array}
$$

and

$$
\mathrm{Cov}[(u_i, v_i, \varepsilon_{2i})'] = \begin{pmatrix} 1 & & \\ .9 & 1 & \\ .5 & .4 & 2 \end{pmatrix}.
$$

In the third case, there is one common variable in both equations which is endogenous:

$$
\begin{array}{lllllll}
y_i^* & = .2 & +.4\,X_{1i} & & +.9\,C_i & & +u_i \\
z_i^* & = 1 & & +.7\,W_{1i} & +.3\,C_i & & +v_i \\
C_i & = .5 & +1.5\,X_{1i} & -.2\,W_{1i} & & +.7\,Z_{3i} & +\varepsilon_{3i}
\end{array}
$$

and

$$\text{Cov}[(u_i, v_i, \varepsilon_{3i})'] = \begin{pmatrix} 1 & & \\ .9 & 1 & \\ .5 & .4 & 2 \end{pmatrix}.$$

Finally, in the fourth case it is assumed that both equations include an endogenous variable which is exclusive for each equation:

$$\begin{array}{lllllllll} y_i^* & = .2 & +.4\,X_{1i} & +.9\,X_{2i} & & & & +u_i \\ z_i^* & = 1 & +.7\,X_{1i} & & +.3\,W_{2i} & & & +v_i \\ X_{2i} & = .5 & +1.5\,X_{1i} & & & +.7\,Z_{1i} & & +\varepsilon_{1i} \\ W_{2i} & = -2 & +1.8\,X_{1i} & & & & +.6\,Z_{2i} & +\varepsilon_{2i} \end{array}$$

and

$$\text{Cov}[(u_i, v_i, \varepsilon_{1i}, \varepsilon_{2i})'] = \begin{pmatrix} 1 & & & \\ .9 & 1 & & \\ .5 & .4 & 2 & \\ .4 & .5 & 1 & 2 \end{pmatrix}.$$

Throughout, X_{1i}, Z_{1i}, Z_{2i} and Z_{3i}, $i = 1, \ldots, n$, are scalars which have been simulated from a standard normal distribution. For each of the four cases, these random numbers have been drawn once and kept fixed during simulation. In total, each simulation encompasses 1000 repetitions in which parameter estimates have been computed. Table 1 presents the mean of these estimates over the repetitions, along with the corresponding standard deviations.

In order to gauge the finite-sample performance of the estimator outlined in section 3, table 6 contains simulation results for different sample sizes. For each sample size, table 6 displays the results for the FIML estimator presented in section 3 ("IV") and contrasts these results with those obtained when using the ordinary estimator for the sample selection model which does not account for endogeneity ("non-IV"). To save space, only the estimates for the parameters of the primary equation and selection equation are presented.

In specification (i) where there is only one endogenous variable included in the primary equation, the IV estimator performs well with respect to the estimates of the primary equation, even for $n = 100$. However, the estimates for the selection equation are upward biased in finite samples; this property is common in all specifications (i)-(iv). In specification (ii) where there is only one endogenous variable in the selection equation, the estimator for the primary equation does well for $n \geq 200$. This is also true for specification (iii) with a common endogenous variable in both equations. When each equation contains an exclusive endogenous variable (specification (iv)), good results are obtained for $n \geq 500$.

Note that the estimates for the selection equation are subjected to a normalization rule. This is the reason why the performance of the IV estimator seems to be not "perfect". However, as it is well known, in binary choice models only coefficient ratios are identified. Put differently, one should not consider the raw coefficients given in table 1 but rather coefficient ratios. For example, in specification (iii) for $n = 1000$ we can calculate that the mean of the second coefficient divided by the first gives 0.7018, whereas the mean of the third coefficient divided by the first gives 0.2991. Thus, we see that also the parameters of the selection equation are well estimated by the FIML procedure.

On the contrary, in most cases the non-IV estimator yields severely biased estimates of the parameters of the primary equation among all specifications. For instance, for a sample size of $n = 1000$ the bias ranges from 13 to 248.1 percent. However, the estimates of the selection equation are sometimes relatively close to their true values (specifications (i) and (iii)). This notwithstanding, note especially that the estimates of the parameters of the main equation are severely biased even if endogeneity is only present in the selection equation (specification (ii)). This result, which is due to the nonlinearity of the underlying model, has not gained much attention in the literature yet.

Overall, the results show that the FIML-IV estimator from section 3 outperforms the ordinary estimator for the sample selection model, especially with respect to the parameters in the primary equation and in case of large sample sizes. Moreover, the results indicate that the bias in the parameter estimates may be substantial if one does not account for endogeneity.

Appendix D

In this appendix, we present an application of our FIML estimator to the labor supply data set introduced by Thomas Mroz (1987). Our goal is to compare our results with those of Wooldridge (2010), who also applied his estimator to this data set.

The Mroz data set is quite popular and is often used to illustrate the performance of estimators which account for sample selectivity. The data set consists of 753 married women of whom 428 are working. We not only have information about relevant labor market characteristics of women (such as the wage, educational attainment and experience) but also on private characteristics such as the number of children, the "non-wife income" and the educational attainment of the parents and the husband. The former variables help identify the selection equation, while the latter variables may serve as instrumental variables for education. These variables are assumed to satisfy an exclusion restriction in the sense that they *directly* affect only the probability of labor market participation and educational attainment, respectively, but not the wage rate.

For this data set, we estimated a wage equation for married women. However, as a wage equation can only be fitted to the subsample of women who are actually working, a simple regression with the women's wage as the dependent variable may yield inconsistent parameter estimates due to the possibility of sample selection. Hence, the appropriate model to estimate the wage equation should be a sample selection model. A variable which is commonly included as an explanatory variable is education. However, there might be some background variables like ability which cannot be observed and, thus, are captured within the error terms. These variables are likely to affect not only wages and labor force participation, but education as well. Therefore, *a priori* education should not be regarded as exogenous. The consequences of falsely treating an endogenous variable like education as exogenous have been illustrated in the preceding section; hence, estimates from the ordinary sample selection model may be severely biased.

We estimated the following model: The main equation contains the natural logarithm of the hourly wage as its dependent variable; explanatory variables are experience, experience squared and education. The selection equation includes experience, experience squared, non-wife income, age, number of children aged until 6 years of age in the household, number of children aged 6 years or older in the household and education. Since education is treated as endogenous, instrumental variables are needed for estimation. Following Wooldridge (2010), we chose mother's education, father's education and husband's education as instrumental variables for education.[9] Means and standard deviations of these variables are presented in table 7.

Estimation results are given in table 8. In table 8, estimation results for the ordinary sample selection model ("non-IV") and the sample selection model with endogeneity ("IV") are provided. The first part of this table contains the parameter estimates for the variables of the main equation, as well as estimates of the "reduced form" selection parameter $\tilde{\rho}$ and the endogeneity parameter ψ_{11}. This last parameter indicates whether endogeneity of education is relevant in the primary equation. The second part presents the parameter estimates for the selection equation. Additionally included is the endogeneity parameter ψ_{21}, which indicates whether endogeneity of education is relevant in the selection equation. Finally, the third part includes the parameter estimates of the exogenous variables and instrumental variables with respect to education. In analogy with the instrumental variables terminology, this part has been labeled "first stage".

The results show significance of education in the primary and the selection equation. Moreover, the instrumental variables for education employed in the "first stage" are highly

[9]For the appropriateness of these instrumental variables, cf. the discussion in Card (1999), pp. 1822-26.

significant. The remaining variables possess the expected signs. However, the estimates of $\tilde{\rho}$, ψ_{11} and ψ_{21} are not significantly different from zero, indicating that there is neither a selection bias nor an endogeneity bias present.[10] These results are in line with those reported by Wooldridge (2010) who draws similar conclusions. However, given that there seems to be neither a sample selection bias nor an endogeneity bias present, this result is not surprising.

[10]In addition, joint significance of ψ_{11} and ψ_{21} is rejected as well (p-value of 0.1907).

Table 1: Descriptive Statistics

	Census 1980		ACS 2005-2010	
	Mean	Std.dev.	Mean	Std.dev.
wage	12.177	3.999	15.306	6.610
educ	12.078	2.480	13.598	2.385
age	38.064	8.965	40.889	8.584
northeast	0.219	0.413	0.178	0.383
midwest	0.268	0.443	0.249	0.433
west	0.188	0.390	0.201	0.401
south	0.326	0.469	0.371	0.483
married	0.786	0.410	0.684	0.465
widowed	0.024	0.153	0.015	0.120
divorced	0.091	0.288	0.132	0.339
separated	0.022	0.147	0.022	0.147
never_married	0.076	0.265	0.147	0.354
nchlt5	0.284	0.590	0.228	0.543
qtr1	0.248	0.432	0.242	0.428
qtr2	0.240	0.427	0.242	0.428
qtr3	0.263	0.440	0.264	0.441
qtr4	0.249	0.432	0.252	0.434
# obs. with nonmissing wage	413,447		948,735	
# obs.	1,032,668		1,528,735	

Table 2: Reduced form estimates for education

	Census 1980		ACS 2005-2010	
	Coeff.	(Std.err)	Coeff.	(Std.err)
age	-0.0316	(0.0028)	-0.0042	(0.0024)
age2	-0.0003	(0.0000)	-0.0003	(0.0000)
northeast	-0.1362	(0.0075)	0.4002	(0.0062)
midwest	-0.1600	(0.0072)	0.1317	(0.0057)
south	-0.4315	(0.0069)	0.0632	(0.0053)
widowed	-0.6909	(0.0157)	-0.8144	(0.0161)
divorced	-0.0247	(0.0083)	-0.3117	(0.0058)
separated	-0.9399	(0.0161)	-1.0684	(0.0130)
never_married	0.4462	(0.0093)	0.1279	(0.0057)
qtr2	-0.0131	(0.0068)	0.0440	(0.0055)
qtr3	0.0432	(0.0066)	0.0304	(0.0054)
qtr4	0.0612	(0.0067)	0.0291	(0.0054)
const	13.8892	(0.0530)	14.1160	(0.0463)
F-statistic	54.0800		22.8200	

Table 3: Mean of education by year and quarter of birth

year/quarter of birth	1	2	3	4
1980	12.05483	12.03125	12.10005	12.12346
2005	13.46413	13.52121	13.50251	13.49254
2006	13.49859	13.53202	13.52017	13.53109
2007	13.54718	13.56337	13.5662	13.55667
2008	13.63228	13.68901	13.65922	13.65476
2009	13.64724	13.6988	13.66024	13.66176
2010	13.6481	13.7246	13.68966	13.68822

Table 4: Heckman estimates without controlling for endogeneity

	Census 1980		ACS 2005-2010	
	Coeff.	(Std.err)	Coeff.	(Std.err)
main equation				
educ	0.0490	(0.0003)	0.1007	(0.0003)
age	0.0191	(0.0006)	0.0470	(0.0005)
age2	0.0002	(0.0000)	0.0005	(0.0000)
northeast	0.0341	(0.0016)	0.0105	(0.0013)
midwest	0.0336	(0.0015)	0.0935	(0.0012)
south	0.0959	(0.0014)	0.0882	(0.0011)
widowed	0.0043	(0.0032)	0.0662	(0.0035)
divorced	0.0258	(0.0020)	0.0108	(0.0013)
separated	0.0132	(0.0031)	0.0769	(0.0028)
never_married	0.0501	(0.0021)	0.0037	(0.0012)
const	1.4512	(0.0119)	0.1550	(0.0106)
selection equation				
educ	0.0959	(0.0006)	0.1439	(0.0005)
age	-0.0822	(0.0016)	-0.0290	(0.0013)
age2	0.0008	(0.0000)	0.0002	(0.0000)
northeast	-0.0695	(0.0043)	0.1083	(0.0035)
midwest	0.0650	(0.0041)	0.2797	(0.0032)
south	0.1432	(0.0039)	0.0817	(0.0030)
widowed	0.3480	(0.0084)	-0.0423	(0.0088)
divorced	1.0032	(0.0048)	0.4387	(0.0034)
separated	0.4970	(0.0088)	0.1100	(0.0072)
never_married	0.8469	(0.0056)	0.2799	(0.0035)
nchlt5	-0.8228	(0.0034)	-0.5529	(0.0023)
const	0.4040	(0.0324)	-0.9434	(0.0276)
correlation parameter	-0.0161	(0.0080)	0.2776	(0.0061)

Table 5: Heckman estimates with controlling for endogeneity

	Census 1980		ACS 2005-2010	
	Coeff.	(Std.err)	Coeff.	(Std.err)
main equation				
educ	0.1078	(0.0225)	0.2309	(0.0436)
age	0.0209	(0.0011)	0.0476	(0.0007)
age2	-0.0002	(0.0000)	-0.0005	(0.0000)
northeast	-0.0261	(0.0034)	-0.0627	(0.0176)
midwest	-0.0242	(0.0043)	-0.1107	(0.0064)
south	-0.0706	(0.0098)	-0.0964	(0.0034)
widowed	0.0363	(0.0162)	0.0399	(0.0361)
divorced	0.0273	(0.0025)	0.0514	(0.0136)
separated	0.0421	(0.0211)	0.0623	(0.0472)
never_married	0.0238	(0.0105)	-0.0129	(0.0061)
eps	-0.0588	(0.0224)	-0.1303	(0.0436)
const	0.6326	(0.3139)	-1.6872	(0.6146)
selection equation				
educ	0.4337	(0.0669)	0.3167	(0.1124)
age	-0.0716	(0.0027)	-0.0283	(0.0018)
age2	0.0009	(0.0000)	0.0003	(0.0000)
northeast	-0.0234	(0.0113)	0.0392	(0.0461)
midwest	0.1190	(0.0128)	0.2570	(0.0161)
south	0.2888	(0.0297)	0.0708	(0.0082)
widowed	0.5814	(0.0462)	0.0984	(0.0923)
divorced	1.0115	(0.0062)	0.4926	(0.0345)
separated	0.8145	(0.0637)	0.2946	(0.1194)
never_married	0.6962	(0.0311)	0.2578	(0.0159)
nchlt5	-0.8228	(0.0043)	-0.5529	(0.0027)
eps	-0.3379	(0.0668)	-0.1728	(0.1124)
const	-4.2966	(0.9243)	-3.3866	(1.5908)
correlation parameter	0.2517	(0.0937)	0.4369	(0.1171)

Note: Standard errors are based on 100 bootstrap replications.

Table 6: Monte Carlo results

Spec.	Param.	$n = 100$ IV	$n = 100$ non-IV	$n = 200$ IV	$n = 200$ non-IV	$n = 500$ IV	$n = 500$ non-IV	$n = 1000$ IV	$n = 1000$ non-IV
(i)	$\beta_1 = .2$	.2397 (.1500)	.1409 (.1498)	.2031 (.0968)	.0934 (.0887)	.2028 (.0556)	.1168 (.0529)	.2014 (.0416)	.0988 (.0381)
	$\beta_2 = .4$	.4019 (.2439)	$-.0191$ (.1535)	.3947 (.1532)	.0396 (.0983)	.4023 (.0945)	.0338 (.0664)	.3988 (.0621)	.0379 (.0413)
	$\beta_3 = .9$	.8991 (.1396)	1.1570 (.0781)	.9020 (.0933)	1.1412 (.0525)	.8978 (.0567)	1.1415 (.0347)	.9007 (.0381)	1.1404 (.0220)
	$\gamma_1 = 1$	1.1316 (.2492)	1.0201 (.1993)	1.1043 (.1467)	1.0101 (.1270)	1.1016 (.0867)	1.0086 (.0758)	1.0995 (.0625)	1.0087 (.0553)
	$\gamma_2 = .7$	.8567 (.2445)	.7483 (.2169)	.7895 (.1337)	.7067 (.1264)	.7724 (.0815)	.6744 (.0795)	.7688 (.0574)	.6707 (.0564)
(ii)	$\beta_1 = .2$	.3068 (.2070)	.6661 (.2250)	.2234 (.1203)	.6784 (.1531)	.2000 (.0597)	.6719 (.1178)	.2001 (.0395)	.6962 (.0642)
	$\beta_2 = .4$	.3082 (.1726)	.0520 (.1892)	.3818 (.1170)	.0181 (.1426)	.4009 (.0561)	.0340 (.1012)	.4000 (.0411)	.0128 (.0584)
	$\gamma_1 = 1$	1.1567 (.2989)	.9346 (.2554)	1.1254 (.1853)	.8766 (.1623)	1.1021 (.1085)	.8544 (.1093)	1.0967 (.0743)	.8541 (.0690)
	$\gamma_2 = .7$	.8226 (.5229)	.2775 (.3628)	.7896 (.3142)	.2177 (.2517)	.7743 (.1624)	.2391 (.1646)	.7708 (.1143)	.2292 (.0994)
	$\gamma_3 = .3$	.3685 (.3325)	.6418 (.2152)	.3451 (.1895)	.6291 (.1403)	.3316 (.0897)	.5854 (.0826)	.3250 (.0672)	.5851 (.0513)
(iii)	$\beta_1 = .2$	.2681 (.1695)	.1575 (.1742)	.2113 (.0987)	.0981 (.1015)	.2010 (.0588)	.0825 (.0570)	.2005 (.0431)	.0863 (.0392)
	$\beta_2 = .4$	.3874 (.2270)	.0147 (.1553)	.4091 (.1554)	.0145 (.1031)	.4007 (.0963)	.0327 (.0631)	.4012 (.0635)	.0348 (.0440)
	$\beta_3 = .9$	.8858 (.1339)	1.1484 (.0829)	.8893 (.0957)	1.1739 (.0588)	.8992 (.0592)	1.1724 (.0346)	.8977 (.0403)	1.1664 (.0238)
	$\gamma_1 = 1$	1.1446 (.2707)	1.0109 (.2044)	1.1222 (.1637)	.9984 (.1346)	1.1044 (.0969)	.9923 (.0861)	1.0987 (.0630)	.9819 (.0561)
	$\gamma_2 = .7$	.8557 (.2600)	.7658 (.2334)	.8053 (.1556)	.7422 (.1520)	.7760 (.0877)	.7292 (.0872)	.7711 (.0582)	.7180 (.0576)
	$\gamma_3 = .3$	.3569 (.1622)	.4696 (.1385)	.3380 (.0834)	.4160 (.0756)	.3324 (.0501)	.4256 (.0455)	.3286 (.0349)	.4216 (.0313)
(iv)	$\beta_1 = .2$	.4320 (.3394)	.3423 (.2752)	.2554 (.2044)	.2899 (.1967)	.1995 (.0835)	.2248 (.0876)	.1988 (.0601)	.2260 (.0649)
	$\beta_2 = .4$	.2738 (.3803)	.0267 (.2147)	.3687 (.2173)	.0735 (.1532)	.4053 (.1219)	.1103 (.0819)	.3994 (.0818)	.1036 (.0603)
	$\beta_3 = .9$	.8887 (.1856)	1.0489 (.0747)	.8965 (.1063)	1.0462 (.0480	.8983 (.0651)	1.0516 (.0304)	.9010 (.0429)	1.0514 (.0209)
	$\gamma_1 = 1$	1.2063 (.5953)	1.5246 (.39175)	1.1415 (.4180)	1.5172 (.2665)	1.0920 (.2316)	1.4562 (.1525)	1.0882 (.1597)	1.4517 (.1111)
	$\gamma_2 = .7$	.8397 (.5378)	.4488 (.2963)	.7793 (.3654)	.4218 (.1890)	.7665 (.2137)	.4216 (.1099)	.7599 (.1391)	.4254 (.0805)
	$\gamma_3 = .3$	.3724 (.2849)	.5504 (.1572)	.3450 (.1935)	.5326 (.1060)	.3281 (.1062)	.5056 (.0604)	.3278 (.0719)	.5041 (.0426)

Table 7: Descriptive statistics for the Mroz data

Variable	Mean	Std.dev.
log wage	4.1777	3.3103
exper	10.6308	8.0691
educ	12.2869	2.2802
nwifeinc	20.1290	11.6348
age	42.5379	8.0726
kidslt6	0.2377	0.5240
kidsge6	1.3533	1.3199
motheduc	9.2510	3.3675
fatheduc	8.8088	3.5723
huseduc	12.4914	3.0208
Sample size	753	
No. of obs. with wage>0	428	

Table 8: Estimation of a wage equation for married women based on the Mroz data

	non-IV		IV	
Main Equation				
const	-0.5527^{**}	(0.2604)	-0.2786	(0.3139)
exper	0.0428^{***}	(0.0149)	0.0449^{***}	(0.0151)
expersq	-0.00008^{**}	(0.0004)	-0.0009^{**}	(0.0004)
educ	0.1084^{***}	(0.0149)	0.0849^{***}	(0.0218)
$\tilde{\rho}$	0.0141	(0.1491)	0.0248	(0.1492)
ψ_{11}			0.0413	(0.0290)
Selection Equation				
const	0.2664	(0.5090)	0.6084	(0.6522)
exper	0.1233^{***}	(0.0187)	0.1261^{***}	(0.0191)
expersq	-0.0019^{***}	(0.0006)	-0.0019^{***}	(0.0006)
nwifeinc	-0.0121^{**}	(0.0049)	-0.0105^{*}	(0.0053)
age	-0.0528^{***}	(0.0085)	-0.0543^{***}	(0.0087)
kidslt6	-0.8674^{***}	(0.1187)	-0.8620^{***}	(0.1190)
kidsge6	0.0359	(0.0435)	0.0316	(0.0438)
educ	0.1313^{***}	(0.0254)	0.1046^{**}	(0.0406)
ψ_{21}			0.0425	(0.0502)
"First Stage"				
const			5.3947^{***}	(0.5826)
exper			0.0577^{***}	(0.0219)
expersq			-0.0008	(0.0007)
nwifeinc			0.0147^{**}	(0.0058)
age			-0.0051	(0.0098)
kidslt6			0.1269	(0.1298)
kidsge6			-0.0700	(0.0511)
motheduc			0.1307^{***}	(0.0224)
fatheduc			0.0951^{***}	(0.0212)
huseduc			0.3489^{***}	(0.0233)

*, ** and *** indicate significance at 1%, 5% and 10%, respectively. Standard errors in parentheses.