Estimating Panel Data Models in the Presence of Endogeneity and Selection

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

Jeffrey M. Wooldridge
Department of Economics
Michigan State University
East Lansing, MI 48824-1038
wooldri1@msu.edu

This version: May 2, 2008

Abstract

We consider estimation of panel data models with sample selection when the equation of interest contains endogenous explanatory variables as well as unobserved heterogeneity. We offer a detailed analysis of the pooled two-stage least squares (pooled 2SLS) and fixed effects-2SLS (FE-2SLS) estimators and discuss complications in correcting for selection biases that arise when instruments are correlated with the unobserved effect. Assuming that appropriate instruments are available, we propose several tests for selection bias and two estimation procedures that correct for selection in the presence of endogenous regressors. The first correction procedure is valid under the assumption that the errors in the selection equation are normally distributed, while the second procedure drops the normality assumption and estimates the model parameters semiparametrically. In the proposed testing and correction procedures, the error terms may be heterogeneously distributed and serially dependent in both selection and primary equations. Correlation between the unobserved effects and explanatory and instrumental variables is permitted. To illustrate and study the performance of the proposed methods, we apply them to estimating earnings equations for females using the Panel Study of Income Dynamics data and perform Monte Carlo simulations.

Keywords: Fixed Effects, Instrumental Variables, Sample Selection, Mills Ratio, Semiparametric

1 Introduction

Due to the increased availability of longitudinal data and recent theoretical advances, panel data models have become widely used in applied work in economics. Common panel data methods account for unobserved heterogeneity characterizing economic agents, something not easily done with pure cross-sectional data.

In many applications of panel data, particularly when the cross-sectional unit is a person, family, or firm, the panel data set is unbalanced. That is, the number of time periods differs by cross-sectional unit. Standard methods such as fixed effects and random effects are easily modified to allow unbalanced panels, but simply implementing the algebraic modifications begs an important question: Why is the panel unbalanced? If the missing time periods result from self-selection, applying standard methods may result in inconsistent estimation.

A number of studies have addressed the problems of heterogeneity and selectivity under the assumption of strictly exogenous explanatory variables. Verbeek and Nijman (1992) proposed two kinds of tests of selection bias in panel data models. The first kind of tests – simple variable addition tests – rely on the assumption of no correlation between the unobserved effects and explanatory variables. Some of their other tests – Hausman-type tests – do not require this assumption, although no suggestion is made on how one can consistently estimate parameters of the model if the hypothesis of no selection bias is rejected. Wooldridge (1995) proposed test and correction procedures that allow the unobserved effects and explanators be correlated in both the selection and primary equations. Distributional assumptions are specified for the error terms in the selection equation, but not for the errors in the primary equation. The model allows idiosyncratic errors in both equations be serially correlated and heterogeneously distributed.

A semiparametric approach to correcting for selection bias was suggested by Kyriazi-

dou (1997). Both the unobserved effects and selection terms are removed by taking the difference between any two periods in which the selection index is the same (or, in practice, "similar"). An important assumption here is that equality of selection indices has the same effect of selection on the dependent variable in the primary equation. Formally, it is assumed that idiosyncratic errors in both equations in the two periods are jointly identically distributed conditional on the explanatory variables and unobserved effects in both equations – a conditional exchangeability assumption. [The conditional exchangeability assumption does not always hold in practice – for example, if variances change over time. Additionally, identification problems may arise when using Kyriazidou's estimator. For a detailed discussion of these issues see Dustmann and Rochina-Barrachina (2007).] Rochina-Barrachina (1999) also uses differencing to eliminate the time-constant unobserved effect; however, in her model the selection is explicitly modeled rather than differenced-out. She assumes trivariate normal distribution of the error terms in the selection and differenced primary equations to derive the selection correction term.

The estimators of Wooldridge (1995), Kyriazidou (1997) and Rochina-Barrachina (1999) help to resolve the endogeneity issues that arise because of non-zero correlation between individual unobserved effects and explanatory variables. However, other endogeneity biases may arise due to a different factor – a nonzero correlation between explanatory variables and idiosyncratic errors. Such type of endogeneity can become an issue due to omission of relevant time-varying factors, simultaneous responses to idiosyncratic shocks, or measurement error. The resulting biases cannot be removed via differencing or fixed effects estimation, and hence, require special consideration.

Extensions to allow for endogenous explanatory variables in the primary equation have been proposed by Vella and Verbeek (1999). In particular, they provide a method for estimating panel data models with censored endogenous regressors and selection, but they do not allow for correlation between the unobserved effects and exogenous variables

in the primary equation. Additionally, when they have more than one endogenous regressor, their approach generally involves multi-dimensional numerical integration, which can be computationally demanding. Kyriazidou (2001) considers estimation of dynamic panel data models with selection. In her model, lags of the dependent variables may appear in both the primary and selection equations, while all other variables are assumed to be strictly exogenous. Charlier, Melenberg, van Soest (2001) show that using instrumental variables (IV) in Kyriazidou's (1997) estimator produces consistent estimators in the presence of endogenous regressors under the appropriate conditional exchangeability assumption, where the conditioning set includes the instruments and unobserved effects in the primary and selection equations. Furthermore, they apply this method to estimating housing expenditure by households. Askildsen, Baltagi, and Holmas (2003) use the same approach when estimating wage elasticity of nurses' labor supply. A somewhat different estimation strategy was proposed by Dustmann and Rochina-Barrachina (2007), who suggest using fitted values in Wooldridge's (1995) estimator, an IV method with generated instruments in Kyriazidou's (1997) estimator, and generalized method of moments (GMM) in Rochina-Barrachina's (1999) estimator. [1] They apply these methods to estimating females' wage equations. Since starting this research, we have come across other extensions of Wooldridge's estimator. Those most closely related to the current work are Gonzalez-Chapela (2004) and Winder (2004). Gonzalez-Chapela uses GMM when estimating the effect of the price of recreation goods on females' labor supply, while Winder uses instrumental variables to account for endogeneity of some regressors when estimating females' earnings equations. Both papers use parametric correction that assumes normality of the error terms in the selection equation. Furthermore, the discussion of the underlying theory in these two papers is quite brief.

As a separate strand of the literature, Lewbel (2005) proposes an estimator that

[1] Additional discussion of how first differencing combined with a double index assumption can be used in the estimation of models with endogenous regressors can be found in Rochina-Barrachina (2000).

addresses endogeneity and selection in panel data models under the assumption that one of explanatory variables is conditionally independent of unobserved heterogeneity and idiosyncratic errors in both primary and selection equations and is conditionally continuously distributed on a large support. The approach employs weighting to address selection and removes fixed effects via differencing. The estimator is a two stage least squares or GMM estimator on the transformed data.

In this study we contribute to the existing literature in several ways. First, we consider two commonly known estimators used in panel data models with endogenous regressors: the pooled two-stage least squares (pooled 2SLS) estimator and fixed effects-2SLS (FE-2SLS) estimator. We show how the presence of unobserved heterogeneity in the selection and primary equation may complicate selection bias correction when the unobserved effect is correlated with exogenous variables. Among other things, our analysis demonstrates that applying cross-sectional correction techniques (such as, for example, the nonparametric estimator of Das, Newey and Vella, 2003) to panel data produces inconsistent estimators, unless one is willing to make a strong assumption that instruments are uncorrelated with (or even independent of) the unobserved heterogeneity.

We propose simple variable addition tests that can be used to detect endogeneity of the sample selection process. These tests, which use functions of the selection indicators from other time periods, can detect correlation between the idiosyncratic error at time t and selection in other time periods. In contrast to Verbeek and Nijman (1992), the proposed tests are robust to the presence of arbitrary correlation between unobserved heterogeneity and explanatory variables. Furthermore, we consider testing for contemporaneous selection bias when enough exogenous variables are observed in every time period. Testing for selection bias is an important first step in analyzing an unbalanced panel because, while one wants to guard against selection bias, selection correction procedures tend to reduce the precision of estimated parameters. Applicability of the tests

described in Verbeek and Nijman (1992) and Wooldridge (1995) is limited because they do not allow for endogenous regressors; they may conclude there is selection bias even if there is none. Our tests are based on the FE-2SLS estimation method, which accounts for endogeneity of regressors in the primary equation, as well as correlated unobserved heterogeneity.

In the case when the test does not reject the hypothesis of no selection bias, we suggest using the FE-2SLS estimator, as it is robust to any type of correlation between unobserved effects and explanatory and instrumental variables, does not require specification of the reduced form equations for endogenous variables, and makes no assumptions of errors distribution.(More efficient GMM estimation is always a possibility, too.) If the hypothesis of no selection bias is rejected, we propose selection correction based on the pooled 2SLS estimator.

We propose two approaches that consider the estimation of population parameters in the presence of endogenous regressors and selection. The first approach is parametric and it uses assumptions that are akin to those specified in Wooldridge (1995). In particular, we assume normality of the errors in the selection equation, and linear conditional mean of the error in the primary equation to derive the correction term. As an alternative approach, we propose a semiparametric estimator that makes no distributional assumptions in the selection and primary equations. Within this approach, the correction term is estimated semiparametrically using series estimators. Both estimators permit heterogeneously distributed and serially dependent errors in the selection equation. Similarly, time heteroskedasticity and arbitrary serial correlation are permitted in the primary equation. Thus, our approach is complementary to Kyriazidou's method (Kyriazidou, 1997) in that our methods allow for arbitrary dynamics in the errors of both equations. Moreover, our semiparametric estimator does not rely on distributional assumptions as in Rochina-Barrachina (1999), and it does not require the availability of a conditionally independent

variable as in Lewbel (2005).

We apply our methods to Panel Study of Income Dynamics (PSID) data, using the years 1980 to 1992. Similarly to Dustmann and Rochina-Barrachina (2007), we estimate earnings equations for females. The finite sample properties of the test and proposed estimators are studied via Monte Carlo simulations.

2 Consistency of Pooled 2SLS

We begin with analyzing the assumptions under which the pooled 2SLS estimator applied to an unbalanced panel is consistent. At this point, we do not explicitly model unobserved heterogeneity, but rather leave it as a part of an error term. Specifically, the main equation of interest is

$$y_{it} = x_{it}\beta + v_{it}, \quad t = 1, \ldots, T \tag{1}$$

where x_{it} is a $1 \times K$ vector that contains both exogenous and endogenous explanatory variables, β is a $K \times 1$ vector of parameters, and v_{it} is the error term. Additionally, assume there exists a $1 \times L$ vector of instruments ($L \geq K$), z_{it}, such that the contemporaneous exogeneity assumption holds for all variables in z_{it}: $\mathrm{E}(v_{it}|z_{it}) = 0$, $t = 1, \ldots, T$. Unless stated otherwise, vectors x_{it} and z_{it} always contain an intercept. Instruments are assumed to be sufficiently partially correlated with the explanatory variables in the population analog of equation (1). In fact, z_{it} includes all the variables in x_{it} that are exogenous in (1). Under the specified assumptions the pooled 2SLS estimator on a balanced panel is consistent.

As a next step, we introduce selection (or incidental truncation) into the model. Let s_{it} be a selection indicator, which equals one if (y_{it}, x_{it}, z_{it}) is observed, and zero otherwise.

Then the pooled 2SLS estimator on an unbalanced panel is

$$\begin{aligned}\hat{\beta}_{2SLS} &= \beta + \left[\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x_{it}'z_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}z_{it}'z_{it}\right)^{-1}\right. \\ &\quad \left.\times\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}z_{it}'x_{it}\right)\right]^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x_{it}'z_{it}\right) \\ &\quad \times\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}z_{it}'z_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}z_{it}'v_{it}\right). \qquad (2)\end{aligned}$$

For fixed T with $N \to \infty$, we can essentially read off conditions that are sufficient for consistency of the pooled 2SLS estimator. These conditions extend those in Wooldridge (2002, Section 17.2.1) for the pure cross sectional case. We summarize with a set of assumptions and a proposition.

ASSUMPTION 2.1: (i) (y_{it}, x_{it}, z_{it}) is observed whenever $s_{it} = 1$; (ii) $\mathrm{E}(v_{it}|z_{it}, s_{it}) = 0$, $t = 1, \ldots, T$; (iii) rank $\mathrm{E}\left(\sum_{t=1}^{T} s_{it}z_{it}'x_{it}\right) = K$; (iv) rank $\mathrm{E}\left(\sum_{t=1}^{T} s_{it}z_{it}'z_{it}\right) = L$.

PROPOSITION 2.1: Under Assumption 2.1 and standard regularity conditions, the pooled 2SLS estimator is consistent and $\sqrt{N}$-asymptotically normal for β.

Assumption 2.1(iv) imposes nonsingularity on the outer product of the instrument matrix in the selected sample. Typically, it is satisfied unless instruments are redundant or the selection mechanism selects too small a subset of the population. Assumption 2.1(iii) is the important rank condition – again, on the selected subpopulation – that requires that we have enough instruments ($L \geq K$) and that they are sufficiently correlated with x_{it}. Any exogenous variable in x_{it} would be included in z_{it}.

Assumption 2.1(ii) is the sense in which selection is assumed to be exogenous in (2).[2]

[2]As is seen from equation (2), a weaker sufficient condition, $\mathrm{E}(s_{it}z_{it}'v_{it}) = 0$, can be used instead of

It requires that v_{it} is conditionally mean independent of z_{it} and selection in time period t. This assumption will be violated if s_{it} is correlated with v_{it}, including cases where v_{it} contains a time-constant unobserved effect that is related to selection. As we will see in Section 5, often an augmented equation will satisfy Assumption 2.1(ii) even when the original population model does not, in which case we can apply pooled 2SLS directly to the augmented equation (provided we have sufficient instruments). Assumption 2.1(ii) is silent on the relationship between v_{it} and s_{ir}, $r \neq t$. In other words, selection is assumed to be contemporaneously exogenous but not strictly exogenous. Consequently, consistency of the pooled 2SLS estimator can hold even if y_{it} reacts to selection in the previous time period, $s_{i,t-1}$, or if selection next period, $s_{i,t+1}$, reacts to unexpected changes in y_{it} (as measured by v_{it}). Of course, if v_{it} contains time-constant unobserved heterogeneity that is correlated with s_{it}, then s_{ir} is likely to be correlated with v_{it}, too. Similarly, if instruments are correlated with omitted unobserved heterogeneity, Assumption 2.1(ii) will fail. Nevertheless, in Section 5 we will put Proposition 2.1 to good use in models with unobserved heterogeneity that is correlated with both instrumental variables and selection.

Importantly, Proposition 2.1 does not impose restrictions on the nature of the endogenous elements of x_{it}. For example, we do not need to assume reduced forms linear in z_{it} with additive, independent, or even zero conditional mean, errors. Consequently, Proposition 2.1 can apply to binary endogenous variables or other variables with discreteness in their distributions. The rank condition Assumption 2.1(iii) can hold quite generally, and is essentially a restriction on the linear projection of x_{it} on z_{it} in the selected subpopulation.

Assumption 2.1(ii). Here, we focus on the conditional mean assumption, as selection correction and tests will be based on that assumption.

3 FE-2SLS and Simple Variable Addition Tests

In many applications of panel data methods, we want to include unobserved heterogeneity in the equation that can be correlated with explanatory variables, and even instrumental variables. In this and subsequent sections we explicitly model the error term as a sum of an unobserved effect and an idiosyncratic error. Therefore, the model is now

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, \ldots, T, \tag{3}$$

where c_i is the unobserved effect and u_{it} are the idiosyncratic errors. We allow for arbitrary correlation between the unobserved effect and explanatory variables. In addition, we allow some elements of x_{it} to be correlated with the idiosyncratic error, u_{it}, as occurs in simultaneous equations models, measurement error, and time-varying omitted variables. In order to allow for correlation between the regressors and the idiosyncratic errors, we assume the existence of instruments, z_{it}, which are strictly exogenous conditional on c_i. This permits for unspecified correlation between z_{it} and c_i, but requires z_{it} to be uncorrelated with $\{u_{ir} : r = 1, ..., T\}$. The dimensions of x_{it} and z_{it} are the same as in the previous section, but, since the FE estimator involves time-demeaning, we assume that all variables in x_{it} and z_{it} are time-varying.

We want to determine assumptions under which ignoring selection will result in a consistent estimator. For each i and t, define $\ddot{x}_{it} \equiv x_{it} - T_i^{-1} \sum_{r=1}^{T} s_{ir} x_{ir}$, where $T_i =$

$\sum_{t=1}^{T} s_{it}$, and similarly for $\ddot{z}_{it}$, $\ddot{y}_{it}$. Then the FE-2SLS estimator can be written as

$$\begin{aligned}\hat{\beta}_{FE-2SLS} &= \left[\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}_{it}'\ddot{z}_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{z}_{it}\right)^{-1}\right. \\ &\times \left.\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{x}_{it}\right)\right]^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}_{it}'\ddot{z}_{it}\right) \\ &\times \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{z}_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{y}_{it}\right), \qquad (4)\end{aligned}$$

which, using straightforward algebra, can be shown to be equal to

$$\begin{aligned}\beta &+ \left[\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}_{it}'\ddot{z}_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{z}_{it}\right)^{-1}\right. \\ &\times \left.\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{x}_{it}\right)\right]^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}_{it}'\ddot{z}_{it}\right) \\ &\times \left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{z}_{it}\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'u_{it}\right). \qquad (5)\end{aligned}$$

The benefit of the within transformation is that it removes the unobserved effect, c_i. Of course, it also means that we cannot estimate coefficients on any time-constant explanatory variables.

Denote $z_i = (z_{i1}, \ldots, z_{iT})$ and $s_i = (s_{i1}, \ldots, s_{iT})$. For consistency of the FE-2SLS estimator on an unbalanced panel, we make the following assumptions:

ASSUMPTION 3.1: (i) (y_{it}, x_{it}, z_{it}) is observed whenever $s_{it} = 1$; (ii) $\mathrm{E}(u_{it}|z_i, s_i, c_i) = 0$, $t = 1, \ldots, T$; (iii) rank $\mathrm{E}\left(\sum_{t=1}^{T} s_{it}\ddot{x}_{it}'\ddot{z}_{it}\right) = K$; (iv) rank $\mathrm{E}\left(\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{z}_{it}\right) = L$.

PROPOSITION 3.1: Under Assumption 3.1 and standard regularity conditions, the FE-2SLS estimator is consistent and $\sqrt{N}$-asymptotically normal for β.

Assuming we have sufficient time-varying instruments, Assumption 3.1(ii) is the critical assumption. By iterated expectations, 3.1(ii) guarantees that $\mathrm{E}\left(\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'u_{it}\right) = 0$. Thus, the last term in equation (5) converges to zero in probability as $N \to \infty$.

Assumption 3.1(ii) always holds if the z_{it} are strictly exogenous, conditional on c_i, and the s_{it} are completely random – so that s_i is independent of (u_{it}, z_i, c_i) in all periods. It also holds when s_{it} is a deterministic function of (z_i, c_i) for all t. In either case we have $\mathrm{E}(u_{it}|z_i, s_i, c_i) = \mathrm{E}(u_{it}|z_i, c_i) = 0, t = 1, \ldots, T$. Allowing for arbitrary correlation between s_{it} and c_i is why fixed effects methods are attractive for unbalanced panels when one suspects different propensities to attrit or otherwise select out of the sample based on unobserved heterogeneity. Random effects (RE) estimation would require, in addition to 3.1(ii), $\mathrm{E}(c_i|z_i, s_i) = 0$, and so RE is not preferred to fixed effects unless selection is truly exogenous.

Allowing for arbitrary correlation between s_{it} and c_i does come at a price. In particular, Assumption 3.1(ii) is not strictly weaker than Assumption 2.1(ii) because 3.1(ii) requires that u_{it} is uncorrelated with selection indicators in all time periods. If we apply Assumption 2.1(ii) to the current context, the pooled 2SLS estimator is consistent if $\mathrm{E}(c_i + u_{it}|z_{it}, s_{it}) = 0$. Granted, with the presence of c_i, it is unlikely that 2.1(ii) would hold when 3.1(ii) does not. But, without an unobserved effect – for example, in a model with a lagged dependent variable and no unobserved effect – 2.1(ii) becomes much more plausible than 3.1(ii). The distinctions between these two assumptions will surface again in Section 5.

Inference for the FE-2SLS estimator on the unbalanced panel can be carried out using standard statistics or, even better, statistics that are robust to heteroskedasticity and serial correlation in $\{u_{it} : t = 1, ..., T\}$. See Wooldridge (1995) for the case of strictly exogenous regressors; the arguments are very similar.

Assumption 3.1 suggests some simple variable addition tests for selection bias. Because Assumption 3.1(ii) implies that u_{it} is uncorrelated with s_{ir} for all t and r, we can add time-varying functions of the selection indicators as explanatory variables and obtain simple t or joint Wald tests. For example, we can add $s_{i,t-1}$ or $s_{i,t+1}$ to (3) and test their significance; we lose a time period (either the first or last) in doing so. Two other possibilities are $\sum_{r=1}^{t-1} s_{ir}$ (the number of times in the sample prior to time period t) and $\sum_{r=t+1}^{T} s_{ir}$ (the number of times in the sample after time period t). For cases of attrition, where attrition is an absorbing state, neither $s_{i,t-1}$ or $\sum_{r=1}^{t-1} s_{ir}$ varies across i for the selected sample, so they cannot be used to test for attrition bias. But $s_{i,t+1}$ and $\sum_{r=t+1}^{T} s_{ir}$ can be used to test for attrition bias.

Adding functions of the selection indicators from other time periods is simple and should have power for detecting selection mechanisms that cause inconsistency in the FE-2SLS estimator. Insofar as the selection indicators are correlated over time, the tests described here will have some ability to detect contemporaneous selection. However, correlation between s_{it} and u_{it} cannot be directly tested by adding selection indicators in an auxiliary regression: it never makes sense to add s_{it} at time t because, by definition, $s_{it} = 1$ for all t in the selected sample. The next section allows us to test for contemporaneous correlation between u_{it} and s_{it} if the set of exogenous instrumental variables is observed in each time period.

4 Testing for Selection Bias Under Incidental Truncation

One way to test for contemporaneous selection bias is to model $\mathrm{E}(v_{it}|z_{it}, s_{it})$ in equation (1). We could then estimate the equation with the additional term inserted and test for selection using the t-test or the Wald test. This type of test has been proposed by Verbeek and Nijman (1992) for panel data models with exogenous explanatory variables. However, if v_{it} includes an unobserved effect, we might conclude there is selection bias simply because the unobserved effect is correlated with some explanatory variables. Here, we build on the test proposed by Wooldridge (1995), which tests for selection bias after estimation by fixed effects. In particular, we extend this approach to allow the possibility that some explanatory variables are not strictly exogenous even after we remove the unobserved effect.

Because fixed effects methods allow selection to be correlated with unobserved heterogeneity, it has advantages over random effects methods. Our approach here is to assume that, in the absence of evidence to the contrary, a researcher applies fixed effects 2SLS to an unbalanced panel. The goal is to then test whether there is sample selection correlated with the idiosyncratic error in the primary equation.

To accommodate specific models of selection, we change the notation slightly from the previous section and write the primary equation as

$$y_{it1} = x_{it}\beta_1 + c_{i1} + u_{it1}, \quad t = 1, \ldots, T, \tag{6}$$

where x_{it} is a $1 \times K$ vector of explanatory variables (some of which can be endogenous), β_1 is a $K \times 1$ vector of parameters, c_{i1} is the unobserved effect and u_{it1} is the idiosyncratic error. Let z_{it} still denote a $1 \times L$ vector of instruments, which are strictly exogenous

conditional on c_{i1}. It is assumed that both x_{it} and z_{it} contain an intercept. In most panel data models, different time intercepts are usually implicit. Unlike in the previous section we now assume that the instrumental variables z_{it} are always observed, while (y_{it1}, x_{it1}) are only observed when the selection indicator, now denoted s_{it2}, is unity. To obtain a test it is convenient to define a latent variable, s^*_{it2},

$$s^*_{it2} = z_{it}\delta_2 + c_{i2} + u_{it2}, \quad t = 1, \ldots, T. \tag{7}$$

Here c_{i2} is an unobserved effect and u_{it2} is an idiosyncratic error. The selection indicator, s_{it2}, is generated as

$$s_{it2} = 1[s^*_{it2} > 0] = 1[z_{it}\delta_2 + c_{i2} + u_{it2} > 0], \tag{8}$$

where $1[\cdot]$ is the indicator function. We will derive a test under the assumption

$$u_{it2}|z_i, c_{i2} \sim \text{Normal}(0, 1), \quad t = 1, \ldots, T, \tag{9}$$

so that s_{it2} follows an unobserved effects probit model. We allow arbitrary serial dependence in $\{u_{it2}\}$.

To proceed further, we model the relationship between the unobserved effect, c_{i2}, and the strictly exogenous variables, z_i. We use the modeling device as in Mundlak (1978). In particular, assume that the unobserved effect can be modeled as

$$c_{i2} = \bar{z}_i\xi_2 + a_{i2}, \tag{10}$$

$$a_{i2}|z_i \sim \text{Normal}(0, \tau_2^2), \quad t = 1, \ldots, T, \tag{11}$$

which assumes that the correlation between c_{i2} and z_i acts only through the time averages of the exogenous variables, while the remaining part of the unobserved effect, a_{i2}, is

independent of z_i. Less restrictive specifications for c_{i2} are possible. A popular option is to assume that $\mathrm{E}(c_{i2}|z_i)$ is a linear projection on z_{i1}, ..., z_{iT}, as in Chamberlain (1980):

$$c_{i2} = z_{i1}\xi_{21} + \ldots + z_{iT}\xi_{2T} + a_{i2}. \tag{12}$$

Mundlak's specification is a special case of Chamberlain's in that (10) imposes the same coefficients ($\xi_{21} = \ldots = \xi_{2T}$) in (12). The advantage of Mundlak's model is that it conserves on degrees of freedom, which is important especially when T is large. In linear panel data models with exogenous explanatory variables and no selection, Mundlak's model produces the estimators of β_1 that are identical to usual fixed effects estimators (Mundlak, 1978). In the case of a binary dependent variable model with normally distributed error terms it leads to a special version of Chamberlain's correlated random effects probit model. In what follows, we use (10).

If we combine (7) through (11) we can write the selection indicator as

$$\begin{aligned} s_{it2} &= 1[z_{it}\delta_2 + \bar{z}_i\xi_2 + v_{it2} > 0] && (13)\\ v_{it2}|z_i &\sim \mathrm{Normal}(0, 1+\tau_2^2), \quad t = 1, \ldots, T, && (14) \end{aligned}$$

where $v_{it2} = a_{i2} + u_{it2}$. In fact, for tests and corrections for selection bias, (13) and (14) are more restrictive than necessary. In many cases, we want to allow coefficients in the selection equations for different time periods to be entirely unrestricted. After all, for the purposes of selection corrections, the selection equation is just a reduced form equation. Therefore, somewhat abusing notation, we specify the following sequence of models:

$$\begin{aligned} s_{it2} &= 1[z_{it}\delta_{t2} + \bar{z}_i\xi_{t2} + v_{it2} > 0] && (15)\\ v_{it2}|z_i &\sim \mathrm{Normal}(0, 1), t = 1, \ldots, T, && (16) \end{aligned}$$

Time varying coefficients on the time average can arise from a standard probit model if we allow the variance of the idiosyncratic term to change over time or if we make the effect of c_{i2} in equation (8) time varying. Typically, there would be some restrictions on the parameters over time, but we will use the flexibility of (15) and (16) because it is more robust.

Given the above (nominal) assumptions and some additional ones, we can derive a test for selection bias. Similar to Wooldridge (1995), suppose (u_{it1}, v_{i2}) is independent of (z_i, c_{i1}), where $v_{i2} = (v_{i12}, \ldots, v_{iT2})'$, and (u_{it1}, v_{it2}) is independent of $(v_{i12}, \ldots, v_{i,t-1,2}, v_{i,t+1,2}, \ldots, v_{iT2})$. Then, if $\mathrm{E}(u_{it1}|v_{it2})$ is linear,

$$\mathrm{E}(u_{it1}|z_i, c_{i1}, v_{i2}) = \mathrm{E}(u_{it1}|v_{i2}) = \mathrm{E}(u_{it1}|v_{it2}) = \rho_1 v_{it2}, \quad t = 1, \ldots, T, \tag{17}$$

where, for now, we assume a regression coefficient, ρ_1, constant across time. Independence of v_{i2} and c_{i1} would not be a good assumption if v_{i2} contains an unobserved effect, as we expect, but, at this point, we are using these assumptions to motivate a test for selection bias. In Section 5 we will be more formal about stating assumptions used for a consistent correction procedure.

From Assumption 3.1 we know that for the FE-2SLS estimator to be consistent on an unbalanced panel, it should be that $\mathrm{E}(u_{it1}|z_i, c_{i1}, s_{i2}) = 0$. If selection is not random, this expectation will depend on the selection indicators and the z_{it}. Under the previous assumptions, we can write

$$\mathrm{E}(u_{it1}|z_i, c_{i1}, s_{i2}) = \rho_1 \mathrm{E}(v_{it2}|z_i, c_{i1}, s_{i2}) = \rho_1 \mathrm{E}(v_{it2}|z_i, s_{it2}), \quad t = 1, ..., T. \tag{18}$$

Now, we can augment the primary equation as

$$y_{it1} = x_{it}\beta + c_{i1} + \rho_1 \mathrm{E}(v_{it2}|z_i, s_{it2}) + e_{it1}, \quad t = 1, ..., T, \tag{19}$$

where, by construction, $\mathrm{E}(e_{it1}|z_i, c_{i1}, s_{i2}) = 0$, $t = 1, ..., T$. It follows that, if we knew $\mathrm{E}(v_{it2}|z_i, s_{it2})$, then a test for selection bias is obtained by testing $H_0 : \rho_1 = 0$ in (19), which we can estimate by FE-2SLS. Of course, since we are only using observations with $s_{it2} = 1$ we need only find $\mathrm{E}(v_{it2}|z_i, s_{it2} = 1)$, and this follows from the usual probit calculation:

$$\mathrm{E}(v_{it2}|z_i, s_{it2} = 1) = \lambda(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}), \quad t = 1, \ldots, T, \tag{20}$$

where $\lambda(\cdot)$ denotes the inverse Mills ratio. Then the following procedure can be used to test for sample selection:

PROCEDURE 4.1 (Valid under the null hypothesis, Assumption 3.1):

(i) For each time period, use the probit model to estimate the equation

$$\mathrm{P}(s_{it2} = 1|z_i) = \Phi(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}). \tag{21}$$

Use the resulting estimates to obtain the inverse Mills ratios, $\hat{\lambda}_{it2} \equiv \lambda(z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2})$.

(ii) For the selected sample, estimate (19) using FE-2SLS, but where $\hat{\lambda}_{it2}$ is in place of $\mathrm{E}(v_{it2}|z_i, s_{it2})$. In addition to $\hat{\lambda}_{it2}$, we can also add the interactions of the inverse Mills ratio with time dummies to allow for different correlations between the idiosyncratic errors u_{it1} and v_{it2} (to allow ρ_1 be different across t).

(iii) Use the t-statistic for ρ_1 to test the hypothesis $H_0 : \rho_1 = 0$, or, in the case when the interactions of the inverse Mills ratio and time dummies are added, use the Wald test to test joint significance of those terms. The variance matrix robust to serial correlation and heteroskedasticity should be used.

If the null hypothesis is true, that is, there is no selection problem, then the FE-2SLS

estimator is consistent, although this particular test only checks for contemporaneous selection. As we discussed in Section 3, the FE-2SLS estimator is consistent even if there is arbitrary correlation between the unobserved effect and the instrumental variables, and it allows selection to be correlated with c_{i1}, too. It does not require us to specify the reduced form equations for the endogenous variables and it imposes no distributional assumptions on u_{it1}. Finally, the serial correlation in u_{it1} is not restricted in any way. Generally, the test in Procedure 4.1 should be useful for detecting selection at time t that is correlated with u_{it1}. The tests in Section 3 can be used to determine if selection in time period t is correlated with the idiosyncratic errors in other time periods – another condition required for consistency of FE-2SLS.

5 Correcting for Selection Bias

5.1 General Setup

If the test described in Section 4 rejects the hypothesis of no selection bias (that depends on the idiosyncratic errors), then a selection correction procedure is needed. As noted earlier, the procedure described in the previous section works for testing, but it can not be used to correct for selection bias. The main problem is the appearance of an unobserved effect inside the index of the probit selection model. If an unobserved effect is present in the selection equation, the error terms in that equation are inevitably serially correlated, which implies a very complicated form for the conditional expectation $\mathrm{E}(v_{it2}|z_i, s_{i2})$. (Plus, some of the other assumptions would be unrealistic, too.) Fortunately, provided we make appropriate linearity assumptions about the conditional expectation of c_{i1}, as in Chamberlain (1980) and Mundlak (1978), we can obtain a valid selection correction.

Specifically, model the unobserved effect as

$$c_{i1} = \bar{z}_i \xi_1 + a_{i1}, \tag{22}$$

$$\mathrm{E}(a_{i1}|z_i) = 0. \tag{23}$$

This condition is akin to (10) and may seem a bit restrictive. However, it in fact is very similar in spirit to the traditional fixed effects estimator. As mentioned earlier, imposing assumptions (22) and (23) in linear panel data models with exogenous explanatory variables ($x_{it} = z_{it}$, $t = 1, \ldots, T$) produces the estimators of slope parameters that are identical to fixed-effects estimators when the estimation is performed on a balanced panel (Mundlak 1978). In equation (22), z_i contains all exogenous variables from the original equation, and hence the effects of those variables in the primary equation are identified off of their deviations from the individual-specific means. With regard to the endogenous variables, their coefficients are identified off of the deviations in the instrumental variables from their within-individual average values. This is very similar to traditional fixed-effects estimation, where the unobserved heterogeneity is assumed to be time-invariant. Naturally, individual-specific time means of exogenous variables vary with T; however, this does not cause a threat to the consistency of the estimator. The asymptotic properties of the considered estimators are for T fixed with $N \to \infty$. Even though the time means are imprecise and change as the time span changes, the corresponding discrepancies go away when averaged across individuals.

Another key feature of condition (22) is that the time means of exogenous variables are obtained on the data that are not distorted by selection (here we exploit the assumption that z_{it} are observed for all i and t). This is one feature that crucially distinguishes the proposed estimator from a standard fixed-effects estimator that performs time-demeaning on a selected sample. While being ideologically similar to fixed effects, the model in (22)

and (23) is free of selection biases, which makes it an attractive modeling device.

Given condition (22) and (23), we can plug into (6) and obtain

$$y_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + a_{i1} + u_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + v_{it1}, \quad t = 1, \ldots, T, \tag{24}$$

where $v_{it1} \equiv a_{i1} + u_{it1}$ and is mean-independent of z_i in the balanced panel. Once we introduce selection that is correlated with unobserved heterogeneity and idiosyncratic errors in the primary equation, it is useful to write

$$y_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + \mathrm{E}(v_{it1}|z_i, s_{it2}) + e_{it1}, \tag{25}$$

$$\mathrm{E}(e_{it1}|z_i, s_{it2}) = 0, \quad t = 1, \ldots, T. \tag{26}$$

So, if we know $\mathrm{E}(v_{it1}|z_i, s_{it2})$, the consistency of the pooled 2SLS estimator would follow by Proposition 2.1.

Note how we do not assert that $\mathrm{E}(e_{it1}|z_i, s_{i2}) = 0$; in fact, generally e_{it1} will be correlated with selection indicators s_{ir2} for $r \neq t$. This is an important benefit of the current approach: we can ignore selection in other time periods that might be correlated with u_{it}. Equations (25) and (26) also show that applying the Mundlak-Chamberlain device to the unbalanced panel, even without a selection term, can be consistent even when the fixed effects estimator is not. Recall that for consistency of the FE-2SLS estimator on the unbalanced sample, selection must be strictly exogenous conditional on c_{i1}. It is plausible that v_{it1} and v_{it2} might be uncorrelated – so $\mathrm{E}(v_{it1}|z_i, s_{it2}) = 0$ – even though $s_{i,t-1,2}$ is correlated with u_{it1}. If so, FE-2SLS is generally inconsistent but adding $\bar{z}_i$ in each time period and using pooled 2SLS is consistent. (Of course, this assumes we observe z_{it} in every time period.)

Generally, equations (25) and (26) show how we can correct for selection by applying pooled 2SLS to (25), at least once we find $\mathrm{E}(v_{it1}|z_i, s_{it2})$. It is possible to make paramet-

ric assumptions and find the exact expression for $\mathrm{E}(v_{it1}|z_i, s_{it2})$, or use semiparametric methods. We consider both approaches below.

5.2 Parametric Correction

A formal set of assumptions that allow us to derive the correction term in parametric setting is as follows.

ASSUMPTION 5.2.1: (i) z_{it} is always observed while (x_{it1}, y_{it1}) is observed when $s_{it2} = 1$; (ii) Selection occurs according to equations (15) and (16); (iii) c_{i1} satisfies (22) and (23); (iv) $\mathrm{E}(v_{it1}|z_i, v_{it2}) \equiv \mathrm{E}(u_{it1}+a_{i1}|z_i, v_{it2}) = \mathrm{E}(u_{it1}+a_{i1}|v_{it2}) = \gamma_{t1}v_{it2}, t = 1, \ldots, T$.

From parts (iii) and (iv) of Assumption 5.2.1 it follows that

$$y_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + \gamma_{t1}\mathrm{E}(v_{it2}|z_i, s_{it2}) + e_{it1} \tag{27}$$

$$\mathrm{E}(e_{it1}|z_i, s_{it2}) = 0, \quad t = 1, ..., T. \tag{28}$$

Conditioning on the selection indicator in the above equation is necessary, as we do not observe v_{it2}. It also suggests that we need to find $\mathrm{E}(v_{it2}|z_i, s_{it2})$ to be able to correct for selection. We already derived this expectation in the previous section, at least for $s_{it2} = 1$ (which is all we need). With a slight abuse of notation, it is convenient to think of writing the equation for $s_{it2} = 1$:

$$y_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + \gamma_{t1}\lambda_{it2} + e_{it1}, \quad t = 1, \ldots, T. \tag{29}$$

This means we can estimate β_1, ξ_1, and $(\gamma_{11}, ..., \gamma_{T1})$ by pooled 2SLS once we replace λ_{it2} (the inverse Mills ratio) with $\hat{\lambda}_{it2}$. We summarize the method for estimating β_1 with the following procedure:

PROCEDURE 5.2.1:

(i) For each time period, run probit of s_{it2} on $1, z_{it}, \bar{z}_i, i = 1, \ldots, N$, and obtain the inverse Mills ratios, $\hat{\lambda}_{it2}$.

(ii) For the selected sample, estimate equation (29) (with λ_{it2} replaced by $\hat{\lambda}_{it2}$) by pooled 2SLS using $1, z_{it}, \bar{z}_i, \hat{\lambda}_{it2}$ as instruments. Note that (29) implies different coefficients for λ_{it2} in each time period. As before, this can be implemented by adding the appropriate interaction terms in the regression. Alternatively, one may estimate a restricted model with $\gamma_{t1} = \gamma_1$ for all t.

(iii) Estimate the asymptotic variance as described in Appendix A.

Instead of using analytical formulae for the asymptotic variance, one can apply "panel bootstrap." This involves resampling cross-sectional units (and all time periods for each unit sampled) and using the bootstrap sample to approximate the distribution of the parameter vector. Such a bootstrap estimator will be consistent for $N \to \infty$ and T fixed.

Moreover, to perform Procedure 5.2.1, we should have a sufficient number of instruments. In particular, if there are Q endogenous variables in x_{it1}, then z_{it} should contain at least $Q + 1$ exogenous elements that are not also in x_{it1}. Effectively, we should have at least one instrument for each endogenous variable, plus at least one additional instrument that affects selection. If we do not have an additional variable that has some separate effect on selection, then the parameters in equation (29) are identified only because of the nonlinearity of the inverse Mills ratio. Often, λ_{it2} will be well approximated by a linear function of most of its range, and the resulting collinearity if λ_{it2} does not depend on a separate variable can lead to very large standard errors.

5.3 Semiparametric Correction

In this section, we relax the assumption of normally distributed errors in the selection equation and propose a semiparametric estimator that is robust to a wide variety of actual error distributions.

As demonstrated below, semiparametric correction permits identification of parameters in β_1 only in the presence of an exclusion restriction. To emphasize this condition formally, we define a vector of instruments used for estimating the primary equation, z_{it1}, $t = 1, \ldots, T$, where z_{it1} has dimension $1 \times L_1$, with $K \leq L_1 < L$. We maintain the assumption that all exogenous elements of x_{it} are included in the set of instruments and also assume that all elements of z_{it1} are included in z_{it} (i.e. z_{it1} is a subset of z_{it}). Because the intercept is not identified when estimating the model semiparametrically, the constant is excluded from the vectors of explanatory and instrumental variables.

To derive the estimating equation, we formulate the following assumptions.

ASSUMPTION 5.3.1: (i) z_{it} is always observed while (x_{it1}, y_{it1}) is observed when $s_{it2} = 1$; (ii) Selection occurs according to equation (15); (iii) c_{i1} satisfies (22), so that the primary equation is given by (24); (iv) The distribution of (v_{it1}, v_{it2}) is either independent of z_i or is a function of selection index $(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2})$.

Notice that Assumption 5.3.1 does not specify a particular form of error distribution, which makes the resulting estimator robust to variations in the distribution of (v_{it1}, v_{it2}). Moreover, it leaves us agnostic about the relationship between the error terms in different time periods, thus, permitting serial correlation, as well as arbitrary relationships between v_{it1} and v_{is2} for $s \neq t$. Part (iv) of Assumption 5.3.1, albeit somewhat restrictive, is routinely used in the literature on semiparametric estimation (Powell, 1994).

From parts (ii) and (iv) of Assumption 5.3.1 it follows that

$$\mathrm{E}(v_{it1}|z_i, s_{it2} = 1) = \varphi_t(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}) \equiv \varphi_{it}, \tag{30}$$

where $\varphi_t(\cdot)$ is an unknown function that may be different in each time period. Hence, combining equations (25) and (30), we can write for $s_{it2} = 1$:

$$y_{it1} = x_{it1}\beta_1 + \bar{z}_i\xi_1 + \varphi_{it} + e_{it1}, \quad t = 1, \ldots, T. \tag{31}$$

To estimate equation (31), we use an approach similar to the one proposed by Newey (1988) and employ series estimators to approximate the unknown function $\varphi_t(\cdot)$. Specifically, the focus is on power series and splines – estimators that are commonly used in economic applications. These are the polynomial and piecewise polynomial functions of the selection index, respectively, and can be easily implemented in practice. In case of splines, the attention is limited to splines with fixed evenly spaced knots.

For estimation purposes it may be preferred to limit the size of the selection index, which in the case of the power series estimator can be done by applying a strictly monotonic transformation $\tau_{it} \equiv \tau(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2})$. Several simple possibilities proposed by Newey (1988, 1994) are logit transformation ($\tau_{it} = [1 + \exp(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2})]^{-1}$), standard normal transformation ($\tau_{it} = \Phi(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2})$), and the inverse Mills ratio. Such a transformation will not alter consistency of the estimator, but will reduce both the effect of outliers and multicollinearity in the approximating terms (Newey, 1994). Similarly, B-splines can be used in place of usual splines to avoid the multicollinearity problem.

Define the vector of M approximating functions as

$$p_{it} \equiv p(\tau_{it}) = (p_1(\tau_{it}), p_2(\tau_{it}), \ldots, p_M(\tau_{it})). \tag{32}$$

Assuming that consistent estimators of δ_{t2} and ξ_{t2} (and hence, τ_{it}) are available, an estimator of β_1 can be obtained by applying pooled 2SLS to equation (31), where φ_{it} is replaced with a linear combination of approximating functions $p(\hat{\tau}_{it})$, $\hat{\tau}_{it} \equiv \tau(z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2})$.

Before formulating consistency assumptions, it is convenient to write the estimator explicitly. Define vectors $w_{it} = (x_{it1}, \bar{z}_i)$, $h_{it} = (z_{it1}, \bar{z}_i)$, $q_{it} = (z_{it}, \bar{z}_i)$, $\theta = (\beta_1', \xi_1')'$, and $\pi_t = (\delta_{t2}', \xi_{t2}')'$. Also, define linear projections of w_{it} and h_{it} on the approximating functions, $\hat{p}_{it} \equiv p(\hat{\tau}_{it})$:

$$\begin{aligned}
\hat{m}_{it}^w &= \hat{p}_{it}\left(\sum_{i=1}^N s_{it2}\hat{p}_{it}'\hat{p}_{it}\right)^{-1}\left(\sum_{i=1}^N s_{it2}\hat{p}_{it}'w_{it}\right), \\
\hat{m}_{it}^h &= \hat{p}_{it}\left(\sum_{i=1}^N s_{it2}\hat{p}_{it}'\hat{p}_{it}\right)^{-1}\left(\sum_{i=1}^N s_{it2}\hat{p}_{it}'h_{it}\right), \quad t = 1, \ldots, T.
\end{aligned} \tag{33}$$

Using the results for partial regression, the estimator of θ can be written as

$$\begin{aligned}
\hat{\theta} \;=\; & \left\{\sum_{t=1}^T\sum_{i=1}^N s_{it2}(w_{it} - \hat{m}_{it}^w)'h_{it}\left(\sum_{t=1}^T\sum_{i=1}^N s_{it2}(h_{it} - \hat{m}_{it}^h)'h_{it}\right)^{-1}\right. \\
\times \; & \left.\sum_{t=1}^T\sum_{i=1}^N s_{it2}(h_{it} - \hat{m}_{it}^h)'w_{it}\right\}^{-1}\sum_{t=1}^T\sum_{i=1}^N s_{it2}(w_{its} - \hat{m}_{it}^w)'h_{it} \\
\times \; & \left(\sum_{t=1}^T\sum_{i=1}^N s_{it2}(h_{it} - \hat{m}_{it}^h)'h_{it}\right)^{-1}\sum_{t=1}^T\sum_{i=1}^N s_{it2}(h_{it} - \hat{m}_{it}^h)'y_{it1}.
\end{aligned} \tag{34}$$

Notice that linear projections $\hat{m}_{it}^w$ and $\hat{m}_{it}^h$ are semiparametric estimators of conditional means, $m_t^w \equiv \mathrm{E}(w_{it}|q_{it}\pi_t, s_{it2} = 1)$ and $m_t^{h1} \equiv \mathrm{E}(h_{it}|q_{it}\pi_t, s_{it2} = 1)$, respectively. In other words, the estimator can be obtained by removing the selection effect via "demeaning," and then applying pooled 2SLS estimator to the transformed data. In this sense, the estimator in (34) is similar to Robinson's estimator (Robinson, 1988).

Given the expression in (34), we can specify the identification assumption:

ASSUMPTION 5.3.2: (i) For $A \equiv \sum_{t=1}^{T} \mathrm{E}\left[s_{it2}(w_{it} - m_t^w)'(h_{it} - m_t^h)\right]$, $\mathrm{rank}(A) = K + L$; (ii) For $B \equiv \sum_{t=1}^{T} \mathrm{E}\left[s_{it2}(h_{it} - m_t^h)'(h_{it} - m_t^h)\right]$, $\mathrm{rank}(B) = L_1 + L$; (iii) For $\Omega \equiv \mathrm{E}\left[\left(\sum_{t=1}^{T} s_{it2}(h_{it} - m_t^h)'e_{it1}\right)\left(\sum_{t=1}^{T} s_{it2}e_{it1}(h_{it} - m_t^h)\right)\right]$, $\mathrm{rank}(\Omega) = L_1 + L$.

Assumption 5.3.2 imposes certain restrictions on the instruments and explanatory variables. In particular, it implies that the number of explanatory variables in the selection equation should be strictly greater than the number of instruments used in the estimation of the primary equation. If this is not the case, "demeaned" instruments may be perfectly linearly related, so that matrices A, B and Ω will not have full rank. The usual requirement that demeaned instruments are sufficiently correlated with demeaned endogenous variables applies.

The following regularity conditions are the same as or similar to those stated in Newey (1988).

ASSUMPTION 5.3.3: (i) $\mathrm{E}(s_{it2}\|w_{it}\|^{2+\nu}) < \infty$ for some $\nu > 0$, $t = 1, \ldots, T$, where the Euclidean norm is defined as $\|C\| = [\mathrm{tr}(C'C)]^{1/2}$; (ii) $\mathrm{E}(s_{it2}\|h_{it}\|^2) < \infty$ for $t = 1, \ldots, T$; (iii) $\mathrm{Var}(w_{it}|q_{it}\pi_t, s_{it2} = 1)$ is bounded for $t = 1, \ldots, T$; (iv) $\mathrm{Var}(h_{it}|q_{it}\pi_t, s_{it2} = 1)$ is bounded for $t = 1, \ldots, T$; (v) $\mathrm{E}(e_{it1}^2|q_{it}\pi_t, s_{it2} = 1)$ is bounded for $t = 1, \ldots, T$.

Assumption 5.3.3 imposes restrictions on conditional and unconditional moments of the variables. These conditions permit the use of the law of large numbers and central limit theorem, as well as secure that series approximations lead to the consistent estimation of the approximated functions.

We further assume that a semiparametric estimator of π_t is available and satifies the following assumption:

ASSUMPTION 5.3.4: For some ψ_{it}, $\sqrt{N}(\hat{\pi}_t - \pi_t) = N^{-1/2}\sum_{i=1}^{N}\psi_{it} + o_p(1) \xrightarrow{d} Normal(0, V_{t2})$, and there exists an estimator $\hat{V}_{t2}$, such that $\hat{V}_{t2} \xrightarrow{p} V_{t2} = \mathrm{E}(\psi_{it}\psi_{it}')$ for $t = 1, \ldots, T$.

Assumption 5.3.4 states that the first-step semiparametric estimator can be approximated as a sample average and is $\sqrt{N}$-consistent and asymptotically normal. Such estimators exist and are described in the literature, the estimators of Ichimura (1993) and Klein and Spady (1993) being the well-known examples. Hence, the first-step estimation should not cause any serious problems, at least in theory.

The last assumption defines properties of φ_t, conditional variable means, and approximating functions, and is very similar to the assumptions formulated by Newey (1988).

ASSUMPTION 5.3.5: (i) Functions φ_t, m_t^w, and m_t^h are continuously differentiable in their argument of orders d, d_w and d_h, respectively, for $t = 1, \ldots, T$; (ii) The distribution of $\tau(q_{it}\pi_t)$ has an absolutely continuous component with p.d.f. bounded away from zero on its support, which is compact. The first and second derivatives of $\tau(q_{it}\hat{\pi}_t)$ with respect to the selection index are bounded for $\hat{\pi}_t$ in a neighborhood of π_t. All variables in q_{it} are bounded; (iii) $M \to \infty$, $N \to \infty$ so that $\sqrt{N}M^{-d-d_h+1} \to 0$ and (a) $p(\tau)$ is a power series, $d \geq 5$, and $M^7/N \to 0$; or (b) $p(\tau)$ is a spline of degree l, with $l \geq d_h - 1$, $d \geq 3$, and $M^4/N \to 0$.

Smoothness conditions in part (i) of Assumption 5.3.5 control for the bias when functions φ_t, m_t^w, and m_t^h are approximated by power series or splines. These conditions, combined with parts (iii)-(v) of Assumption 5.3.3, guarantee that $\hat{\varphi}_t$, $\hat{m}_t^w$, and $\hat{m}_t^h$ converge in probability to their true values as the number of approximating terms grows. Similarly, additional smoothness requirements in part (iii) of Assumption 5.3.5 are neces-

sary to ensure consistent estimation of θ and the first derivative of φ_t. These smoothness assumptions are not restrictive and are commonly used in the literature. Moreover, as noted by Newey (1988), and Donald and Newey (1994), from part (iii) of Assumption 5.3.5 it appears that $\sqrt{N}$-consistency of θ does not require undersmoothing, i.e. the number of the approximating terms need not grow faster than the optimum in order to reduce the approximation bias of $\hat{\varphi}_t$. If m_t^h is smooth enough, then undersmoothing for φ_t is not necessary. Part (ii) of Assumption 5.3.5 imposes restrictions on the transformation function and the variables in the selection equation. Boundedness of τ_{it} and h_{it} is not restrictive in practice, while the requirement for τ_{it} to have p.d.f. which is bounded away from zero is somewhat restrictive. Both conditions are needed, however, for series approximations to work.

PROPOSITION 5.3.1: Under Assumptions 5.3.1-5.3.5, $\hat{\theta}$ is consistent and $\sqrt{N}$ - asymptotically normal for θ.

In summary, $\hat{\theta}$ can be obtained by implementing the following two-step procedure:

PROCEDURE 5.3.1:

(i) For each time period, use a semiparametric estimator that satisfies Assumption 5.3.4 to obtain $\hat{\pi}_t$, $t = 1, \ldots, T$; compute $p(\hat{\tau}_{it})$.

(ii) For the selected sample, estimate equation (31) (with φ_{it} replaced by the set of approximating functions) by pooled 2SLS using $z_{it1}, \bar{z}_i, p(\hat{\tau}_{it})$ as instruments. One can allow the selection correction to be different in each time period by adding the appropriate interaction terms in the regression.

(iii) Estimate the asymptotic variance as described in Appendix A.

6 Empirical Application

The estimation and testing procedures described above can be used in a variety of settings. Here we estimate a wage offer equation for females, similar to the analysis in Dustmann and Rochina-Barrachina (2007). The main goal is to obtain estimates for the return to labor force experience.

As discussed in the literature, longitudinal earnings equations for females are likely to suffer from heterogeneity, endogeneity, and selection biases. Heterogeneity is usually associated with individual ability and motivation. Since these factors are likely to be correlated with at least some explanatory variables (for instance, education), simple estimation methods, such as pooled OLS, will not produce consistent estimators. Endogeneity of experience is another potential problem. Apart from the fact that experience can be correlated with ability, if the participation decision in each period depends on the wage offer, then an exogenous shock to wages in the past will be correlated with the number of years of experience we observe today. Thus, experience cannot be regarded as strictly exogenous even after conditioning on the unobserved effect. Finally, selection is a potential problem because we observe the wage offer only for women who choose to work, and participation is possibly correlated with idiosyncratic changes in the wage offer.

These three problems can be tackled by applying estimation and testing methods discussed in the previous sections of this paper. At this point, we offer a word of caution about how one implements the selection methods in Sections 4 and 5. It is important to implement the methods as described, avoiding temptation to generate fitted values from a first-stage estimation and then plug those fitted values into the primary equation before correcting for selection bias. To see why, suppose that the model includes only one endogenous explanatory variable, x_{itk}, which is always observed. Then we can think of using the entire sample to estimate a linear reduced form,

$$x_{itk} = \eta_1 z_{it1} + ... + \eta_L z_{itL} + b_i + r_{it}, \quad t = 1, ..., T, \tag{35}$$

where b_i is an unobserved effect. However, estimating (35) by fixed effects and plugging the fitted values, say $\hat{x}_{itk}$, in for x_{it}, is tantamount to replacing x_{itk} with $\eta_1 z_{it1} + ... + \eta_L z_{itL} + b_i$ and then putting r_{it} as part of the idiosyncratic error. In other words, first estimating this equation by fixed effects and substituting in the fitted values is tantamount to applying selection correction to the composite term $r_{it} + v_{it1}$, rather than just v_{it1}. This may be legitimate, but for certain kinds of endogenous variables x_{itk}, the assumptions used in deriving the correction term will fail, thus invalidating the correction procedure. For example, if x_{itk} is binary, then r_{it} is the error in a linear probability model, and $\mathrm{E}(r_{it}|v_{it2})$ is definitely not linear. Consequently, $\mathrm{E}(r_{it} + v_{it1}|v_{it2})$ will be nonlinear and part (iv) of Assumption 5.2.1 will fail. Applying pooled 2SLS directly to (29) or (31) is the most robust procedure because it does not take a stand on the nature of x_{itk}, and, therefore, it does not impose strong restrictions on its reduced form.

Plugging in fitted values is even more problematical when some endogenous explanatory variables are nonlinear functions of other endogenous variables. Typical wage equations – and ours is no exception – include experience as a quadratic (or some more complicated polynomial). The way to handle such equations is to view any function of an endogenous variable as just another endogenous variable, in which case we need to find instruments for these nonlinear functions. To be clear on this point, assume there is no sample selection problem, and consider a model that contains a single endogenous explanatory variable, x_{itk}, in level form and through the known, nonlinear, function $g(\cdot)$:

$$y_{it1} = \beta_1 x_{it1} + ... + \beta_{k-1} x_{it,k-1} + \beta_k x_{itk} + \beta_{k+1} g(x_{itk}) + c_i + u_{it}, \quad t = 1, ..., T. \tag{36}$$

Using our approach, we would simply define $x_{it,k+1} \equiv g(x_{itk})$ and then ignore the fact

that $x_{it,k+1}$ is a known function of x_{itk}. Naturally, our choice of instruments for $x_{it,k+1}$ *would* recognize the functional dependence, but how one uses those instruments would not. For example, if $g(x_{itk}) = x_{itk}^2$, the extra instruments would include the squares of at least some elements of z_{it}, possibly along with cross products.

Plugging in fitted values obtained from (35) has no known asymptotic properties for T fixed and $N \to \infty$ (and they are unlikely to be good). Consider the equation

$$y_{it} = \beta_1 x_{it1} + ... + \beta x_{it,k-1} + \beta_k \hat{x}_{itk} + \beta_{k+1} g(\hat{x}_{itk}) + c_i + error_{it}, \quad t = 1, ..., T, \qquad (37)$$

where $error_{it}$ contains estimation error but also errors that arise from replacing a variable with its linear projection. Plus, by inserting $\hat{x}_{itk}$ into $g(\cdot)$, we are effectively saying that the linear projection operator passes through nonlinear functions. An additional problem now is that $\hat{x}_{itk}$ depends on "estimates" of the b_i. With small T, this introduces an incidental parameters problem, and makes it difficult to derive any asymptotic properties of the estimator. [Even without the incidental parameters problem that arises from estimating the b_i, (37) is an example of a "forbidden" regression. See, for example, Wooldridge (2002, Section 9.5.2).]

We now turn to estimation of a wage offer equation using data from the Panel Study of Income Dynamics (PSID) for the years 1980-1992 (survey years 1980-1993). The sample is limited to white females who were either heads of households or "wives," and who remained in the sample during the considered period. [3] The raw sample consists of 1,716 individuals, and it reduces to 864 individuals after imposing age restrictions, excluding self-employed and agricultural workers, and dropping observations with inconsistent or missing data. In particular, a woman was excluded from the analysis if one of the fol-

[3]These sample restrictions are dictated by the fact that years of actual experience are available only for household heads and "wifes." Also, as is standard in the literature, the correction procedures are obtained under the assumption that z_i are always observed, i.e. there is no attrition.

lowing happened in at least one year during 1980 to 1992: self-reported age exceeded the age constructed using information on the year of birth by more than two years or self-reported age was smaller than constructed age by more than one year (76 observations); the woman was less than 18 or more than 65 years old (346 observations); the woman was self-employed (352) or an agricultural worker (15 observations); experience was missing (17 observations); the woman's age exceeded her experience by less than six years (1 observation); the woman reported positive work hours and zero earnings (11 observations); spouse's weeks of unemployment was missing (21 observations); or the change in years of schooling between 1976 and 1985 was negative and exceeded one year in absolute value (13 observations). In cases when the reported decrease in years of schooling was one year, the minimum of the two reported values was assigned in all periods. The final sample consists of 11,232 observations, out of which 8,254 observations contain information on earnings. When estimating earnings equations we restrict our sample to females who worked in at least two years during 1980-1992. The loss of observations due to this restriction is quite modest (18 observations).

In choosing the set of explanatory and instrumental variables we follow Dustmann and Rochina-Barrachina (2007) fairly closely. Using the notation from Sections 4 and 5, the dependent variable in the main equation of interest, y_{it1}, is the log of real average hourly earnings. The average hourly earnings are defined as a ratio of the individual's annual labor income and annual hours worked; all earnings data were deflated to 1983 dollars using the consumer price index. The vector of explanatory variables, x_{it1}, includes education measured in years of schooling, experience, experience squared, and time dummies. In the PSID, experience is not available for each survey year. We construct this variable by taking the information about prior experience from 1976 survey year or from the year when the individual entered the sample for the first time, and then updating this information annually. In each year experience was increased by one if the annual work

hours were 2000 or more, and it was increased by the number of hours worked divided by 2000 if the annual work hours were less than 2000. Education is considered to be strictly exogenous conditional on the unobserved effect, while experience is not strictly exogenous. The set of instruments, z_{it}, consists of the following variables: years of schooling, time dummies, age and its square, an indicator for marital status, other family income and its square, number of children in the family in three age categories, age of the spouse (who can be either a legal spouse or an important other residing together) and its square, spouse's education and its square, number of weeks the spouse was unemployed, and an indicator for whether the spouse's weeks of unemployment were not reported for various reasons. The selection rule is for labor force participation. A woman is considered to be a participant if she reports positive work hours in a given year. Summary statistics for the variables used in the analysis are presented in Table 1.

As discussed in the previous section, the semiparametric Procedure 5.3.1 will produce consistent estimators of parameters in β_1 only if exclusion restrictions are available. In the considered application, the decision to work or not to work is likely to be affected by spouse's employment status in the current period. On the other hand, current employment status of the spouse should not affect woman's current experience, since experience is determined by past labor-leisure choices. This restriction validates the exclusion of spouse's weeks of unemployment and an indicator of whether this information was not reported from the set of instruments used in the estimation of the primary equation. We impose this exclusion restriction when using the semiparametric estimator.

Table 2 reports the coefficient estimates from seven different estimation methods. Pooled OLS assumes that all explanatory variables are uncorrelated with unobserved heterogeneity and are also strictly exogenous. Pooled 2SLS instruments for the experience variables, but does not remove an unobserved effect. (So, for example, schooling is assumed to be uncorrelated with unobserved ability and motivation.) Fixed effects allows

for correlation between the explanatory variables and unobserved heterogeneity while FE-2SLS further allows experience (and its square, of course) to be correlated with the idiosyncratic errors. Nevertheless, FE-2SLS assumes that selection into the workforce is not systematically related to idiosyncratic changes in the earnings equation.

To determine whether there is evidence of selection bias in using FE-2SLS on the unbalanced panel, we compute two of the statistics described at the end of Section 3. Namely, we add (one at a time) $s_{i,t-1}$ and $s_{i,t+1}$ as explanatory variables in the FE-2SLS estimation. The coefficient on $s_{i,t-1}$ is 0.176 with a fully robust t statistics of 4.89. When we use $s_{i,t+1}$ we get a coefficient of 0.061 with $t = 1.66$. There is strong evidence that wage at time t is higher for those in the labor force in the previous year, and some evidence that next year's participation is positively correlated with wage shocks. In any case, a correction procedure seems warranted.

Strictly speaking, Procedure 4.1 corrects for selection bias – it is FE-2SLS with inverse Mills ratio terms added – only under restrictive assumptions on the selection mechanism. It is presented mainly because the joint Wald test on the 13 Mills ratio terms, made robust to arbitrary serial correlation and heteroskedasticity, provides further evidence of selection bias in using FE-2SLS. The chi-square statistic, with 13 degrees-of-freedom, is 26.96, which gives a p-value of about 0.0126. As with the tests based on selection indicators, there is statistically significant evidence of selection bias.

Procedure 5.2.1 makes distributional assumptions for the errors in the selection equation and corrects for contemporaneous selection while remaining agnostic about whether selection is strictly exogenous. The estimated return to the first year of experience, roughly 5.5%, is notably lower for Procedure 5.2.1 than for the procedures discussed earlier. But the marginal effect declines less sharply than in the other estimation methods. The bottom part of Table 2 shows the return to experience at 12 years of experience (roughly the average in the sample). The estimated return from Procedure 5.2.1, roughly 4.5%, is

equal to the estimate obtained from the FE-2SLS regression. However, the standard error is somewhat larger once selection is accounted for. Procedure 5.2.1 also estimates a larger turning point: the estimated return to experience becomes negative only after 64 years, which far past the highest experience in the sample (roughly 45 years). Thus, according to these estimates, the return to experience never becomes negative over the range of the data and beyond that.

Not surprisingly, the years of schooling estimate is reduced dramatically by controlling for an unobserved effect, but it is still statistically significant.

Before applying semiparametric estimation, we perform a series of specification tests in the selection equation to find out whether the normality assumption may not hold. In each time period, we estimate equation (21) by probit and use resulting estimates to compute fitted values for the selection index. Then, the selection equation was augmented by the second and third powers of the selection index and estimated by probit. We use the standard Wald test to test the hypothesis that additional terms are jointly insignificant, i.e. initial probit model is correct. The hypothesis was rejected at the 10% level in two cases, at the 5% level – in two cases, and at the 1% level – in two cases, suggesting that the parametric assumptions of Section 5.2 may be too strong.

When implementing the semiparametric approach of Section 5.3, we estimate the selection equation by Ichimura's (1993) estimator. Prior to estimation, a linear transformation was applied to the explanatory variables to obtain the sample covariance equal to an identity matrix. The bandwidth was selected using the method proposed by Hardle, Hall and Ichimura (1993), who suggest minimizing the mean squared error simultaneously with respect to both parameters and the bandwidth. The search for the optimal bandwidth was performed on the interval $[0.1, 0.4]$ at 10 grid points. The kernel function was chosen to be Gaussian. In the primary equation, we use the standard normal transformation of the selection index to construct the approximating functions. Linear and quadratic

terms, as well as their interactions with time dummies, were used to approximate φ_{it}. [4]

Estimates from the semiparametric correction procedure are reported in the last column of Table 2. Procedure 5.3 produces the coefficient estimate on the linear experience term of about 4.8%, which is somewhat smaller than the estimate from Procedure 5.2. In other words, the return to the first year of experience reduces further when we use semiparametric correction. The marginal effect of experience evaluated at 12 years is also smaller (only 3.6%). Due to lower estimated returns, Procedure 5.3.1 also gives a smaller turning point (roughly 48 years), although it is still beyond the maximal years of experience in the sample. In summary, correcting for endogeneity of experience and sample selection results in flattening of the earnings-experience profile. Not surprisingly, it also gives larger standard errors.

7 Simulations

In this section we present the results of limited Monte Carlo simulations that demonstrate the properties of the test and estimators in finite samples. We consider a model described by equations (6) and (8), where x_{it} is a scalar, z_{it} is a vector of two variables, and $\beta_1 = \delta_{21} = \delta_{22} = 1$. Unobserved effects, c_{i1} and c_{i2}, are independent across i and distributed as $Normal(0, \sigma_c^2)$. Idiosyncratic errors, u_{it1} and u_{it2}, are independent across i and t and distributed as $Normal(0, \sigma_u^2)$. The total variance of the composite errors, $c_{i1} + u_{it1}$ and $c_{i2} + u_{it2}$, is $\sigma^2 = \sigma_c^2 + \sigma_u^2 = 1$; the proportion of the total variance due to the unobserved effect, σ_c^2/σ_u^2, varies across experiments. The correlation between the

[4] We also tried including third and forth order polynomials of the transformed selection index and the corresponding interactions with time dummies, but the higher power functions turned out to be highly collinear with the linear and quadratic terms. Moreover, including more approximating functions had little influence on the estimated coefficients of education, experience, and experience squared. Therefore, we chose to limit out attention to linear and quadratic terms.

unobserved effects is equal to 0.7, while $\rho_{u_1,u_2} \equiv \text{Corr}(u_{it1}, u_{it2})$ varies depending on the experiment.

The endogenous and exogenous variables were generated as follows:

$$\begin{aligned} z_{it1} &= b_{i1} + \epsilon_{it1}, \\ z_{it2} &= b_{i2} + \epsilon_{it2}, \\ x_{it} &= z_{it1} + \zeta u_{it1} + b_{i3} + \epsilon_{it3}, \end{aligned} \tag{38}$$

where unobserved effects, b_{i1}, b_{i2}, and b_{i3}, are independent across i and distributed as $Normal(0, \sigma_b^2)$; idiosyncratic errors, ϵ_{i1}, ϵ_{i2}, and ϵ_{i3}, are independent across i and t and distributed as $Normal(0, \sigma_\epsilon^2)$. The total variance is $\sigma^2 = \sigma_b^2 + \sigma_\epsilon^2 = 1$, and the proportion of the total variance due to the corresponding unobserved effect changes from experiment to experiment. The correlation between any two unobserved effects (including c_{i1} and c_{i2}) is equal to 0.7. Thus, all variables are correlated with each other through the unobserved effects, whenever the unobserved heterogeneity is present. There is also a non-zero correlation between x_{it} and the idiosyncratic component of z_{it1}. Coefficient ζ varies across experiments. When performing simulations, we use $z_{it} = (z_{it1}, z_{it2})$ as regressors in the selection equation and use z_{it1} as an instrument for x_{it}.

Table 3 presents Monte Carlo results for the size and power of the test described in Section 4. Simulations were performed for $N = 200$ and 500, and $T = 5$ and 10, using 1000 replications. Because selection bias may arise due to unobserved heterogeneity, as well as non-zero correlation between u_{it1} and u_{it2}, the computed size of the test appears on the intersection of the first column and the first row in each panel of the table. In all experiments, the computed size of the test is close to the nominal size. The power of the test increases with N and T, as well as when the correlation between the idiosyncratic errors, ρ_{u_1,u_2}, increases. However, when the proportion of the variance due to unobserved

heterogeneity rises, the power of the test is reduced because the share of σ_u^2 falls.

When evaluating the performance of the estimators discussed in Section 5, we focus on the following cases:

(i) $\sigma_c^2 = \sigma_b^2 = \zeta = \rho_{u_1,u_2} = 0$. That is, there is no unobserved heterogeneity, x_{it} is strictly exogenous, and the idiosyncratic errors in the primary and selection equations are independent.

(ii) $\sigma_c^2 = \sigma_b^2 = 0.5,\ \zeta = \rho_{u_1,u_2} = 0$. Here we introduce unobserved heterogeneity, but maintain the assumption of zero correlation with idiosyncratic errors.

(iii) $\sigma_c^2 = \sigma_b^2 = \zeta = 0.5,\ \rho_{u_1,u_2} = 0$. In addition to unobserved heterogeneity, we introduce endogeneity of x_{it} due to it being correlated with the idiosyncratic error in the primary equation.

(iv) $\sigma_c^2 = \sigma_b^2 = \zeta = \rho_{u_1,u_2} = 0.5$. In this case, we have all three components present: unobserved heterogeneity, endogeneity, and selection due to correlation between the idiosyncratic errors.

(v) $\sigma_c^2 = \sigma_b^2 = \zeta = 0.5,\ \rho_{u_1,u_2} = -0.5$. This is almost like case (iv), but the idiosyncratic errors in the selection and primary equation are negatively correlated.

We run simulations for $N = 200$ and $T = 5$. Number of replications is 1000. [5] Results are reported in Table 4. We computed the bias, average standard error and root mean square error (RMSE) for six estimators: OLS, 2SLS, FE, FE-2SLS, the parametric estimator discussed in Section 5.2, and the semiparametric estimator described in Section 5.3. Because in our simulations the number of regressors is equal to the number of instruments, the 2SLS estimator is the same as the instrumental variables (IV) estimator.

[5] When using the semiparametric estimator described in Section 5.3, we estimate the selection equation by Ichimura's (1993) estimator. To reduce the computational burden, the bandwidth was fixed at $(NT)^{-1/5}$.

Average standard error is the average over the replications of the fully-robust standard error (i.e. standard error robust to serial correlation and heteroskedasticity). In the case of selection correction, we compute standard errors that also account for the first-step estimation.

Results in the top part of Table 4 indicate that in the absence of unobserved heterogeneity, endogeneity and selection ($\sigma_c^2 = \sigma_b^2 = \zeta = \rho_{u_1,u_2} = 0$), all six estimators have very small biases. Standard errors and RMSE are the smallest for OLS and are substantially larger for the procedures that correct for selection. Average standard errors of the estimators in Sections 5.2 and 5.3 (as well as of the other estimators) are very similar to RMSE, which implies that estimating variances as suggested by the asymptotic theory produces rather accurate standard errors in small samples.

Once we introduce unobserved heterogeneity ($\sigma_c^2 = \sigma_b^2 = 0.5$), both OLS and 2SLS estimators appear to be biased. The biases of the other estimators are still negligibly small, but the estimators summarized by Procedures 5.2.1 and 5.3.1 appear to be inferior to FE and FE-2SLS estimators because of the relatively high RMSE. Adding endogeneity ($\sigma_c^2 = \sigma_b^2 = \zeta = 0.5$) causes both FE and OLS to be biased. The bias of the 2SLS estimator is also large due to non-zero correlation between the unobserved heterogeneity and z_{it1}. As expected, when $\rho_{u_1,u_2} = 0$, FE-2SLS is clearly preferred to selection correction procedures because of the smaller bias and RMSE. In the last two cases, where ρ_{u_1,u_2} is different from zero, the estimators discussed in Sections 5.2 and 5.3 perform better than all other estimators. Even though their standard errors are relatively large, the biases remain small, so that applying selection correction procedures produces the smallest RMSE.

8 Conclusion

We have shown how to estimate panel data models in the presence of selection when the primary equation contains endogenous explanatory variables, where endogeneity is conditional on the unobserved effect. These models arise in various economic applications, such as estimation of earnings equations and labor supply models; therefore, the methods discussed in this paper should provide a useful tool for applied economic research. The proposed tests offer robust ways of testing for selection bias in the presence of endogenous regressors. The suggested correction procedures provide an important alternative to some existing methods, as they allow general serial correlation on idiosyncratic errors in the primary and selection equations. Additionally, our semiparametric estimator shares the properties of all semiparametric estimators in the sense that it is robust to a wide variety of error distributions. The results of Monte Carlo simulations show that the estimators perform reasonably well in small samples.

An avenue for further research is in relaxing the single-index assumption for the selection equation. Semiparametric and nonparametric procedures that relax the separability of the unobserved effect from the effects of other variables in binary response panel data models (see, for example, Altonji and Matzkin, 2005), can add to the flexibility of the approach.

Appendix A

In this section, we present the derivation of the asymptotic variance of the estimators discussed in Procedures 5.2.1 and 5.3.1. Let either Assumption 5.2.1 hold so that y_{it1} can be written as in (29), or let Assumption 5.3.1 hold and write y_{it1} as in (31). Technically, we should use equation (25) with the parametric or semiparametric form of $\mathrm{E}(v_{it1}|z_i, s_{it2})$

substituted in, but this expectation disappears for $s_{it2} = 0$, anyway. Therefore, we abuse notation slightly and express y_{it1} as in (29) or (31) for the selected sample.

Define the generated regressors and instruments for time period t as $\hat{w}_{it} = (\ x_{it1},\ \bar{z}_i,\ 0,\ \ldots,\ 0,\ \hat{d}_{it2},\ 0,\ \ldots,\ 0)$ and $\hat{h}_{it} = (z_{it1},\ \bar{z}_i,\ 0,\ \ldots,\ 0,\ \hat{d}_{it2},\ 0,\ \ldots,\ 0)$, respectively, where $\hat{d}_{it2} = \hat{\lambda}_{it2}$ (the inverse Mills ratio) if using Procedure 5.2.1, and $\hat{d}_{it2} = \hat{p}_{it2}$ (the set of approximating functions) if using Procedure 5.3.1. In the primary equation, the parameter vector is $\theta = (\beta_1', \xi_1', \gamma_{11}', \ldots, \gamma_{T1}')'$, where γ_{t1} is a scalar when using parametric correction, and it is an $m \times 1$ vector when using series approximations. In the selection equation, the parameter vectors are $\pi_t = (\delta_{t2}', \xi_{t2}')'$, and $\pi = (\pi_1', \pi_2', \ldots, \pi_T')'$. When we drop the "ˆ" over $\hat{w}_{it}$ and $\hat{h}_{it}$, these are evaluated at the unknown population parameter, π, rather than $\hat{\pi}$.

The pooled 2SLS estimator on the selected sample, after plugging in the first-stage estimates from the selection equations, is

$$\begin{aligned}\hat{\theta} \;=\; & \left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{w}_{it}\right)\right]^{-1} \\ & \times\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'y_{it1}\right); \qquad (39)\end{aligned}$$

since $y_{it1} = w_{it}\theta + e_{it1} = \hat{w}_{it}\theta + (w_{it} - \hat{w}_{it})\theta + e_{it1}$, we have upon substitution

$$
\begin{aligned}
\sqrt{N}(\hat{\theta}-\theta) &= \left[\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1}\right. \\
&\quad \left.\times\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{w}_{it}\right)\right]^{-1} \\
&\quad \times\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1} \\
&\quad \times\left(N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'[(w_{it}-\hat{w}_{it})\theta+e_{it1}]\right) \\
&= (C'D^{-1}C)^{-1}C'D^{-1} \\
&\quad \times\left(N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'[(w_{it}-\hat{w}_{it})\theta+e_{it1}]\right)+o_p(1), \qquad (40)
\end{aligned}
$$

where $C \equiv \mathrm{E}\left(\sum_{t=1}^{T} s_{it2}h_{it}'w_{it}\right)$ and $D \equiv \mathrm{E}\left(\sum_{t=1}^{T} s_{it2}h_{it}'h_{it}\right)$. [Naturally, the representation in (40) assumes regularity conditions, but we suppress those here.] Using an argument similar to Wooldridge (2002, Chapter 6) and $\mathrm{E}(e_{it1}|h_{it}, s_{it2}) = 0$,

$$
\begin{aligned}
&\left(N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'[(w_{it}-\hat{w}_{it})\theta+e_{it1}]\right) \\
&= -\mathrm{E}\left[\sum_{t=1}^{T} s_{it2}h_{it}'(\theta'\nabla_{\pi}w_{it}')\right]\sqrt{N}(\hat{\pi}-\pi) \\
&+N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}h_{it}'e_{it1}+o_p(1), \qquad (41)
\end{aligned}
$$

where $\nabla_{\pi}w_{it}'$ is the Jacobian of w_{it}' with respect to π. Because $\hat{\pi}$ is either a vector of probit maximum likelihood estimators for each t, or a vector of semiparametric estimators

satisfying Assumption 5.3.4, we have

$$\sqrt{N}(\hat{\pi} - \pi) = N^{-1/2} \sum_{i=1}^{N} \psi_i(\pi) + o_p(1), \tag{42}$$

where $\psi_i(\pi)$ depends on the expected Hessians and scores for either the probit log-likelihoods or the first-step semiparametric estimators; more on this below. It follows that

$$\begin{aligned} &\left(N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it2} \hat{h}'_{it} [(w_{it} - \hat{w}_{it})\theta + e_{it1}] \right) \\ &= N^{-1/2} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right] + o_p(1), \end{aligned} \tag{43}$$

where $F = \mathrm{E}\left[\sum_{t=1}^{T} s_{it2} h'_{it} (\theta' \nabla_\pi w'_{it})\right]$. Combining (43) and (40) gives

$$\begin{aligned} &\sqrt{N}(\hat{\theta} - \theta) \\ &= (C'D^{-1}C)^{-1} C'D^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right] \right) \\ &+ o_p(1) \end{aligned} \tag{44}$$

and so

$$\sqrt{N}(\hat{\theta} - \theta) \overset{a}{\sim} \mathrm{Normal}[0, (C'D^{-1}C)^{-1} C'D^{-1} G D^{-1} C (C'D^{-1}C)^{-1}] \tag{45}$$

where

$$G = \mathrm{Var}\left(\sum_{t=1}^{T} s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right) \equiv \mathrm{Var}[g_i(\theta, \pi)].$$

Consistent estimation of $\mathrm{Avar}[\sqrt{N}(\hat{\theta} - \theta)]$ is straightforward by replacing unknown pa-

rameters with their consistent estimators. Consistent estimators of C, D, and G are

$$\hat{C} \equiv N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}'_{it}\hat{w}_{it} \tag{46}$$

$$\hat{D} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}'_{it}\hat{h}_{it} \tag{47}$$

$$\hat{G} = N^{-1}\sum_{i=1}^{N} \hat{g}_i\hat{g}'_i, \tag{48}$$

respectively, where $\hat{g}_i = \sum_{t=1}^{T} s_{it2}\hat{h}'_{it}\hat{e}_{it1} - \hat{F}\hat{\psi}_i$, $\hat{e}_{it1} = y_{it1} - \hat{w}_{it}\hat{\theta}$, and $\hat{F} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}'_{it}(\hat{\theta}'\nabla_\pi \hat{w}'_{it})$. Only $\hat{F}$ and $\hat{\psi}_i$ require some work to compute. For $\hat{F}$ we need to obtain $\nabla_\pi \hat{w}'_{it}$. But, for each (i,t), $\nabla_\pi \hat{w}'_{it}$ is easily seen to be a block matrix with all blocks zero except one. Namely, if we let $q_{it} \equiv (z_{it}, \bar{z}_i)$, then

$$\nabla_\pi w'_{it} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & -q_{it}\mu_{it2} & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \tag{49}$$

where $\mu_{it2} = \lambda_{it2} \cdot (q_{it}\pi_t + \lambda_{it2})$ [the derivative of the inverse Mills ratio, see Wooldridge 2002, p. 522] if use Procedure 5.2.1, and $\mu_{it2} = \nabla_{q_{it}\pi_t} p'_{it}$ if use Procedure 5.3.1. Further, $\theta'\nabla_\pi w'_{it} = (0, \ldots, 0, -\gamma_{t1}q_{it}\lambda_{it2} \cdot (q_{it}\pi_t + \lambda_{it2}), 0, \ldots, 0,)$ for the parametric correction, and $\theta'\nabla_\pi w'_{it} = (0, \ldots, 0, -q_{it}\frac{dp_{it}\gamma_{t1}}{d(q_{it}\pi_t)}, 0, \ldots, 0,)$ if correct for selection semiparametrically. So, if use Procedure 5.2.1,

$$\hat{F} = -N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[0, \ldots, 0, s_{it2} \hat{h}'_{it} \hat{\gamma}_{t1} q_{it} \hat{\lambda}_{it2} \cdot (q_{it} \hat{\pi}_t + \hat{\lambda}_{it2}), 0, \ldots, 0 \right]. \tag{50}$$

And, if use Procedure 5.3.1,

$$\hat{F} = -N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[0, \ldots, 0, s_{it2} \hat{h}'_{it} q_{it} \frac{d\hat{p}_{it} \hat{\gamma}_{t1}}{d(q_{it} \pi_t)}, 0, \ldots, 0 \right]. \tag{51}$$

The expressions for $\hat{\psi}_i$, will depend on the first-step estimator used for obtaining the parameter estimates in the selection equation. In the semiparametric case, the formulae will be different depending on the choice of the first-step semiparametric estimator. In the case of parametric correction summarized by Procedure 5.2.1, the formulae are known. Specifically, from standard results for probit, for each i and t we have vectors

$$\hat{\psi}_{it} = \hat{H}_t^{-1} \{ \Phi(q_{it} \hat{\pi}_t)[1 - \Phi(q_{it} \hat{\pi}_t)] \}^{-1} \phi(q_{it} \hat{\pi}_t) q'_{it} [s_{it2} - \Phi(q_{it} \hat{\pi}_t)], \tag{52}$$

where

$$\hat{H}_t \equiv N^{-1} \sum_{i=1}^{N} \{ \Phi(q_{it} \hat{\pi}_t)[1 - \Phi(q_{it} \hat{\pi}_t)] \}^{-1} [\phi(q_{it} \hat{\pi}_t)]^2 q'_{it} q_{it} \tag{53}$$

is the consistent estimator of minus the expected Hessian, and $\hat{\pi}_t$ is the maximum likelihood estimator from probit of s_{it2} on $q_{it}, i = 1, ..., N$. For each i, we stack the $\hat{\psi}_{it}$ to obtain $\hat{\psi}_i$, which are then used in equation (48).

Appendix B

In this section, we provide the proof of Proposition 5.3.1. This proof relies on the results derived by Newey (1988).

Define vectors of variables as in Section 5.3. Specifically, let $w_{it} = (x_{it1}, \bar{z}_i)$, $h_{it} = (z_{it1}, \bar{z}_i)$, $q_{it} = (z_{it}, \bar{z}_i)$, $\theta = (\beta_1', \xi_1')'$, and $\pi_t = (\delta_{t2}', \xi_{t2}')'$. Rewrite (34) to obtain

$$\begin{aligned}
&\hat{\theta} = \theta \\
&+ \left\{ \sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (w_{it} - \hat{m}_{it}^{w})' h_{it} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (h_{it} - \hat{m}_{it}^{h})' h_{it} \right)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (h_{it} - \hat{m}_{it}^{h})' w_{it} \right\}^{-1} \\
&\times \sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (w_{it} - \hat{m}_{it}^{w})' h_{it} \left(\sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (h_{it} - \hat{m}_{it}^{h})' h_{it} \right)^{-1} \\
&\times \sum_{t=1}^{T} \sum_{i=1}^{N} s_{it2} (h_{it} - \hat{m}_{it}^{h})' (\varphi_{it} + e_{it1}),
\end{aligned} \tag{54}$$

where $\varphi_{it} \equiv \varphi_t(q_{it}\pi_t)$. Consider vector $r_{it} = (w_{it}, h_{it})$ and define

$$\begin{aligned}
m_t^r &\equiv \mathrm{E}(r_{it} | q_{it}\pi_t, s_{it2} = 1), \\
\hat{m}_{it}^r &= \hat{p}_{it} \left(\sum_{i=1}^{N} s_{it2} \hat{p}_{it}' \hat{p}_{it} \right)^{-1} \left(\sum_{i=1}^{N} s_{it2} \hat{p}_{it}' r_{it} \right), \quad t = 1, \ldots, T.
\end{aligned} \tag{55}$$

From parts (i)-(iv) of Assumption 5.3.3 and Assumptions 5.3.4 and 5.3.5, it follows as in Newey (1988), proof of Theorem 1, that

$$N^{-1} \sum_{i=1}^{N} s_{it2} (r_{it} - \hat{m}_{it}^r)' r_{it} \xrightarrow{p} \mathrm{E}\left[s_{it2} (r_t - m_t^r)' (r_t - m_t^r) \right], \quad t = 1, \ldots, T. \tag{56}$$

Therefore,

$$N^{-1}\sum_{i=1}^{N} s_{it2}(w_{it}-\hat{m}_{it}^{w})'h_{it} \xrightarrow{p} \mathrm{E}\left[s_{it2}(w_{it}-m_{t}^{w})'(h_{it}-m_{t}^{h})\right], \quad t=1,\ldots,T,$$
$$N^{-1}\sum_{i=1}^{N} s_{it2}(h_{it}-\hat{m}_{it}^{h})'h_{it} \xrightarrow{p} \mathrm{E}\left[s_{it2}(h_{it}-m_{t}^{h})'(h_{it}-m_{t}^{h})\right], \quad t=1,\ldots,T,$$
$$N^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} s_{it2}(w_{it}-\hat{m}_{it}^{w})'h_{it} \xrightarrow{p} A,$$
$$N^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N} s_{it2}(h_{it}-\hat{m}_{it}^{h})'h_{it} \xrightarrow{p} B. \tag{57}$$

Then, under Assumptions 5.3.2(i) and 5.3.2(ii), rearranging equation (54) will give

$$\sqrt{N}(\hat{\theta}-\theta) = (AB^{-1}A)^{-1}AB^{-1}\frac{1}{\sqrt{N}}\sum_{t=1}^{T}\sum_{i=1}^{N} s_{it2}(h_{it}-\hat{m}_{it}^{h})'(e_{it1}+\varphi_{it}) + o_p(1). \tag{58}$$

Also, from parts (ii), (iv) and (v) of Assumption 5.3.3, Assumptions 5.3.4 and 5.3.5, it follows as in Newey (1988), proof of Theorem 1, that for $t=1,\ldots,T$,

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} s_{it2}(h_{it}-\hat{m}_{it}^{h})'(e_{it1}+\varphi_{it}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[s_{it2}(h_{it}-m_{t}^{h})'e_{it1} - F_t\psi_{it}\right] + o_p(1),$$

where $F_t = \mathrm{E}\left[s_{it2}(h_{it}-m_{t}^{h})'\frac{d\varphi_{it}}{d(q_{it}\pi_t)}q_{it}\right]$. Consequently,

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{t=1}^{T}\sum_{i=1}^{N} s_{it2}(h_{it}-\hat{m}_{it}^{h})'(e_{it1}+\varphi_{it})\\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\sum_{t=1}^{T} s_{it2}(h_{it}-m_{t}^{h})'e_{it1} - \sum_{t=1}^{T} F_t\psi_{it}\right] + o_p(1),\end{aligned} \tag{59}$$

So, by the central limit theorem, $\sqrt{N}(\hat{\theta}-\theta) \xrightarrow{d} Normal(0,V)$, where V is the asymptotic variance of $\hat{\theta}$ that can be consistently estimated using the estimator described in Appendix A. The proof of consistency of $\hat{V}$ is available from the authors upon request.

References

Altonji, J.G. and Matzkin, R.L. 2005, Cross section and panel data estimators for non-separable models with endogenous regressors. Econometrica 73, 1053-1102.

Askildsen, J.E., B.H. Baltagi and T.H. Holmas, 2003, Wage policy in the health care sector: a panel data analysis of nurses' labour supply, Health Economics 12, 705-719.

Chamberlain, G., 1980, Analysis with qualitative data, Review of Economic Studies 47, 225-238.

Charlier, E., B. Melenberg, and A. van Soest, 2001, An analysis of housing expenditure using semiparametric models and panel data, Journal of Econometrics 101, 71-107.

Das, M., W.K. Newey, and F. Vella, 2003, Nonparametric estimation of sample selection models. Review of Economic Studies 70, 33-58.

Donald, S.G. and W.K. Newey, 1994, Series estimation of semilinear models. Journal of Multivariate Analysis 50, 30-40.

Dustmann, C. and M.E. Rochina-Barrachina, 2007, Selection correction in panel data models: An application to the estimation of females' wage equations. Econometrics Journal 10, 263-293.

Hardle, W., P. Hall, and H. Ichimura, 1993, Optimal smoothing in single-index models. Annals of Statistics 21, 157-178.

Ichimura, H., 1993, Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. Journal of Econometrics 58, 71120.

Gonzalez-Chapela, J., 2004, On the price of recreation goods as a determinant of female labor supply. Unpublished manuscript.

Klein, R.L. and R.H. Spady, 1993, An efficient semiparametric estimator for binary response models. Econometrica 61, 387421.

Kyriazidou, E., 1997, Estimation of a panel data sample selection model. Econometrica 65, 1335-1364.

Kyriazidou, E., 2001, Estimation of dynamic panel data sample selection models, Review of Economic Studies 68, 543-572.

Lewbel, A., 2005, Simple endogenous binary choice and selection panel model estimators. Unpublished manuscript, Boston College.

Mundlak, Y., 1978, On the pooling of time series and cross section data, Econometrica 46, 69-85.

Newey, W.K., 1988, Two step series estimation of sample selection models. Unpublished manuscript, MIT (revised version January 1999).

Newey, W.K., 1994, The asymptotic variance of semiparametric estimators. Econometrica 62, 1349-1382.

Newey, W.K., 1997, Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147-168.

Powell, J.L., 1994, Estimation of semiparametric models, in: R.F. Engle and D. McFadden, (Eds.), Handbook of Econometrics, Vol. 4. North Holland, Amsterdam, pp. 2444-2521.

Robinson, P.M., 1988, Root-N-consistent semiparametric regression. Econometrica 56, 931-954.

Rochina-Barrachina, M.E., 1999, A new estimator for panel data sample selection models. Annales d'Economie et de Statistique 55/56, 153-181.

Rochina-Barrachina, M.E., 2000, New semiparametric pairwise difference estimators for panel data sample selection models. The 5th chapter of the thesis dissertation "Panel data sample selection models" at University College London (University of London).

Vella, F. and M. Verbeek, 1999, Two-step estimation of panel data models with censored endogenous variables and selection bias, Journal of Econometrics 90, 239-263.

Verbeek, M. and T. Nijman, 1992, Testing for selectivity bias in panel data models, International Economic Review 33, 681-703.

Winder, K.L, 2004, Reconsidering the motherhood wage penalty. Unpublished manuscript.

Wooldridge, J.M., 1995, Selection corrections for panel data models under conditional mean independence assumptions, Journal of Econometrics 68, 115-132.

Wooldridge, J.M., 2002, Econometric analysis of cross section and panel data. MIT Press: Cambridge, MA.

Table 1: Summary Statistics

Variable Description	Entire Sample	Participants	Non-Participants
Participation (=1 if works)	0.73	1	0
Log of Real Hourly Earnings	—	1.94	—
		(0.62)	
Experience (years)	11.76	12.93	8.51
	(7.76)	(7.58)	(7.31)
Education (years)	12.94	13.13	12.40
	(2.27)	(2.24)	(2.29)
Age (years)	40.91	40.12	43.12
	(10.28)	(9.61)	(11.65)
Married (=1 if married)	0.86	0.84	0.93
Other Household Income (thousands)	34.461	31.167	44.268
	(40.586)	(30.996)	(58.520)
Spouse's Age (years)	37.07	35.21	42.21
	(18.05)	(18.14)	(16.75)
Spouse's Education (years)	11.26	11.04	11.88
	(5.22)	(5.47)	(4.38)
Weeks Spouse Unemployed	0.98	0.95	1.06
	(4.96)	(4.79)	(5.39)
Weeks Unreported (=1 if spouse's unemployment unreported)	0.08	0.06	0.16
Children Aged 0-2	0.14	0.11	0.21
	(0.37)	(0.33)	(0.45)
Children Aged 3-5	0.18	0.16	0.24
	(0.42)	(0.40)	(0.49)
Children Aged 6-17	0.82	0.84	0.77
	(1.01)	(1.01)	(0.99)
Number of Observations	11,232	8,254	3,978

Sample standard deviations are in parentheses below sample averages.

Table 2: Estimates for the Log(Hourly Earnings) Equation

Explanatory Variable	Pooled OLS	Pooled 2SLS	Fixed Effects	FE-2SLS	Procedure 4.1	Procedure 5.2.1	Procedure 5.3.1
Experience	0.067***	0.062***	0.082***	0.067***	0.068**	0.055**	0.0476*
	(0.006)	(0.011)	(0.010)	(0.023)	(0.030)	(0.026)	(0.0246)
$Experience^2$	-0.0014***	-0.00166***	-0.00091***	-0.00091***	-0.00066**	-0.00043	-0.0005
	(0.0002)	(0.0004)	(0.00017)	(0.00032)	(0.00033)	(0.00035)	(0.00033)
Education	0.114***	0.115***	0.023**	0.025**	0.024**	0.035***	0.0355***
	(0.006)	(0.007)	(0.0114)	(0.0115)	(0.0116)	(0.013)	(0.0125)
Wald Test of Joint Significance of the Correction Terms					$\chi^2_{13} = 26.96$ (0.0126)	$\chi^2_{13} = 15.03$ (0.305)	$\chi^2_{26} = 38.29$ (0.0569)
Marginal Effect of Experience	0.034***	0.022***	0.060***	0.045**	0.052**	0.045**	0.0357*
	(0.002)	(0.004)	(0.008)	(0.019)	(0.025)	(0.0218)	(0.0205)
Turning Point	23.9	18.7	45.1	36.8	51.5	64.0	47.5

Year dummy variables are included in each regression; results not reported.
Marginal effects are estimated at experience = 12.
Standard errors robust to serial correlation and heteroskedasticity in parentheses under coefficient estimates; robust p-values are under the test statistics.
The standard errors for Procedures 5.2.1 and 5.3.1 are also corrected for the first-step estimation.
*** = significant at the 1% level; ** = significant at the 5% level; * = significant at the 10% level

Table 3: Computed Size and Power of the Test (Procedure 4.1), $\sigma^2 = 1$, $\zeta = 0.5$

ρ_{u_1,u_2}	$\sigma_c^2 = \sigma_b^2 = 0$	$\sigma_c^2 = \sigma_b^2 = 0.3$	$\sigma_c^2 = \sigma_b^2 = 0.5$	$\sigma_c^2 = \sigma_b^2 = 0.7$
		$N = 200$, $T = 5$		
0.0	0.052	0.060	0.056	0.046
0.1	0.120	0.082	0.081	0.066
0.2	0.244	0.187	0.147	0.125
0.3	0.501	0.340	0.260	0.176
0.4	0.750	0.568	0.408	0.266
0.5	0.904	0.750	0.579	0.419
		$N = 200$, $T = 10$		
0.0	0.053	0.040	0.063	0.039
0.1	0.181	0.134	0.126	0.094
0.2	0.556	0.369	0.292	0.195
0.3	0.866	0.707	0.567	0.369
0.4	0.986	0.901	0.822	0.595
0.5	1.000	0.989	0.946	0.793
		$N = 500$, $T = 5$		
0.0	0.048	0.065	0.050	0.055
0.1	0.154	0.126	0.099	0.104
0.2	0.539	0.359	0.277	0.190
0.3	0.885	0.672	0.552	0.392
0.4	0.984	0.904	0.762	0.573
0.5	1.000	0.978	0.924	0.772
		$N = 500$, $T = 10$		
0.0	0.045	0.054	0.039	0.046
0.1	0.386	0.261	0.216	0.153
0.2	0.902	0.753	0.604	0.423
0.3	0.999	0.983	0.915	0.757
0.4	1.000	1.000	0.992	0.948
0.5	1.000	1.000	1.000	0.996

The table displays the fraction of rejections of the null hypothesis that $\rho_1 = 0$ (see equation 19) out of 1000 replications.
The nominal size of the test is 0.05.

Table 4: Performance of Parametric and Semiparametric Estimators, $\sigma^2 = 1$, $N = 200$, $T = 5$

	OLS	2SLS	FE	FE-2SLS	Procedure 5.2.1	Procedure 5.3.1
			$\sigma_c^2 = \sigma_b^2 = \zeta = \rho_{u_1,u_2} = 0$			
Bias	-0.0005	0.0004	0.0008	0.0021	-0.0014	0.0041
Average std. err.	0.0335	0.0507	0.0427	0.0649	0.0671	0.0683
RMSE	0.0326	0.0499	0.0424	0.0628	0.0686	0.0679
			$\sigma_c^2 = \sigma_b^2 = 0.5,\ \zeta = \rho_{u_1,u_2} = 0$			
Bias	0.1928	0.1657	-0.0009	-0.0008	-0.0015	0.0015
Average std. err.	0.0369	0.0488	0.0389	0.0567	0.0636	0.0669
RMSE	0.1964	0.1729	0.0395	0.0546	0.0626	0.0672
			$\sigma_c^2 = \sigma_b^2 = \zeta = 0.5,\ \rho_{u_1,u_2} = 0$			
Bias	0.3336	0.1608	0.2628	-0.0034	-0.0056	-0.0002
Average std. err.	0.0331	0.0474	0.0363	0.0571	0.0643	0.0669
RMSE	0.3354	0.1679	0.2654	0.0577	0.0648	0.0672
			$\sigma_c^2 = \sigma_b^2 = \zeta = \rho_{u_1,u_2} = 0.5$			
Bias	0.2903	0.0965	0.2359	-0.0686	-0.0026	-0.0042
Average std. err.	0.0338	0.0503	0.0368	0.0604	0.0630	0.0664
RMSE	0.2924	0.1096	0.2389	0.0912	0.0635	0.0686
			$\sigma_c^2 = \sigma_b^2 = \zeta = 0.5,\ \rho_{u_1,u_2} = -0.5$			
Bias	0.3734	0.2252	0.2810	0.0601	-0.0010	0.0039
Average std. err.	0.0326	0.0453	0.0347	0.0529	0.0639	0.0660
RMSE	0.3749	0.2299	0.2831	0.0823	0.0668	0.0671

Monte Carlo results are obtained using 1000 replications.

Averaged standard errors are robust to serial correlation and heteroskedasticity.

Standard errors for Procedures 5.2.1 and 5.3.1 are also corrected for the first-step estimation.