

IZA DP No. 162

Selection Correction in Panel Data Models: An Application to Labour Supply and Wages

Christian Dustmann
María Engracia Rochina-Barrachina

June 2000

Selection Correction in Panel Data Models: An Application to Labour Supply and Wages

Christian Dustmann

Department of Economics, University College London, and IZA, Bonn

María Engracia Rochina-Barrachina

Department of Economics, Universidad de Valencia and University College London

Discussion Paper No. 162

June 2000

IZA

P.O. Box 7240

D-53072 Bonn

Germany

Tel.: +49-228-3894-0

Fax: +49-228-3894-210

Email: iza@iza.org

This Discussion Paper is issued within the framework of IZA's research area *Project Evaluation*. Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent, nonprofit limited liability company (Gesellschaft mit beschränkter Haftung) supported by the Deutsche Post AG. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public. The current research program deals with (1) mobility and flexibility of labor markets, (2) internationalization of labor markets and European integration, (3) the welfare state and labor markets, (4) labor markets in transition, (5) the future of work, (6) project evaluation and (7) general labor economics.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character.

ABSTRACT

Selection Correction in Panel Data Models: An Application to Labour Supply and Wages*

In recent years a number of panel estimators have been suggested for sample selection models, where both the selection equation and the equation of interest contain individual effects which are correlated with the explanatory variables. We review and compare some of these estimators, and apply them to estimating the return to actual labour market experience for females, using a panel of twelve years. All these estimators rely on the assumption of strict exogeneity of regressors in the equation of interest, conditional on individual specific effects and the selection mechanism. This assumption is likely to be violated in many applications. Also, life history variables are often measured with error in survey data sets, because they contain a retrospective component. We show how non-strict exogeneity and measurement error can be taken into account within the estimation methods discussed.

JEL Classification: C33, C34, J3

Keywords: Panel data sample selection estimators, female labour supply, estimation of wage equations

Christian Dustmann
Department of Economics
University College London
Gower Street
London WC1E 6BT
UK
Tel.: +44 (0)171 504 5212
Fax: +44 (0)171 916 2775
Email: c.dustmann@ucl.ac.uk

* We are grateful to Richard Blundell, Arthur van Soest and Frank Windmeijer for useful comments and suggestions. Thanks are also owed to participants at the Primer Encuentro de Economía Aplicada, June 1998, Barcelona, Spain; at the 8th-International Conference on Panel Data, June 1998, Göteborg, Sweden; at the German Socio-Economic Panel Users (GSOEP) Conference, July 1998, Berlin, Germany; at the European Meeting of the Econometrics Society (ESEM), August 1998, Berlin, Germany; at the Royal Economic Society (RES) Annual Conference, March-April 1999, Nottingham, United Kingdom; and at the European Society for Population Economics (ESPE) Conference, June 1999, Torino, Italy. Financial support from the Spanish foundation "Fundación Ramón Areces" to María E. Rochina-Barrachina is gratefully acknowledged. The usual disclaimer applies.

1. Introduction

In many problems of applied econometrics, the equation of interest is only defined for a subset of individuals from the overall population, while the parameters of interest are the parameters that refer to the whole population. Examples are the estimation of wage equations, or hours of work equations, where the dependent variable can only be measured when the individual participates in the labour market. If the sub-population is non-randomly drawn from the overall population, straightforward regression analysis leads to inconsistent parameter estimates. This problem is well known as sample selection bias, and a number of estimators are available which correct for this (see Heckman (1979), or Powell (1994) for an overview).

Another problem is the presence of unobserved heterogeneity in the equation of interest. Economic theory often suggests estimation equations that contain an individual specific effect, which is unobserved, but correlated with the model regressors. Examples are unobserved ability components in wage equations, correlated with wages and education (see Card (1994) for details), or the estimation of Frisch demand functions in the consumption and labour supply literature (see, for instance, Browning, Deaton, and Irish (1985), Blundell and MaCurdy (1999) and MaCurdy (1981)). If unobserved individual specific (and time constant) effects affect the outcome variable, and are correlated with the model regressors, simple regression analysis does not identify the parameters of interest. For the estimation of coefficients on variables which vary over time, panel data provide a solution to this latter problem, and a number of straightforward estimators are available (see Chamberlain (1984), and Hsiao (1986) for overviews).

In many applications, both problems occur simultaneously. If the selection process is time constant, panel estimators solve both problems. But often this is not the case. Recently, some estimators have been proposed which deal with both sources of estimation bias. These estimators require panel data, and produce consistent parameter estimates under various sets of assumptions. We consider three estimators which allow for additive individual specific effects in both the (binary) selection equation and the equation of interest, and, at the same time, allow for the equation of interest being defined for a non-

random sub population. These estimators impose different consistency requirements, some of which may be restrictive in particular applications.

The first estimator we consider has been proposed by Wooldridge (1995). It relies on a full parameterisation of the sample selection mechanism, and requires specifying the functional form of the conditional mean of the individual effects in the equation of interest. It does not impose distributional assumptions about the error terms and the fixed effects in the equation of interest. The second estimator we discuss has been proposed by Kyriazidou (1997). The basic idea of this estimator is to match observations within individuals, which have the same selection effect in two time periods, and to difference out both the individual heterogeneity term, and the selection term. The third estimator has been developed in Rochina-Barrachina (1999). This method adds a distributional assumption for the error term in the equation of interest.

In the first part of the paper, we describe the main features of the three estimators, and point out the conditions under which each of them produces consistent estimates of the parameters of interest. Not many applications of these estimators exist in the literature. In the second part of the paper, we apply the three methods to a typical problem in labour economics. We estimate wage equations for female labour market participants, and try to identify the effect of actual labour market experience on wages. In this application, all the before mentioned problems arise. Female labour market participants are non-randomly drawn from the overall population. Their participation propensity depends on unobservables, which are likely to be correlated with the model regressors. And their productivity depends on unobservables, which are likely to be correlated with the regressors in the main equation.

All three estimators impose the assumption of strict exogeneity of the explanatory variables. In many typical applications, like the one we use as an illustration, this assumption is likely to be violated. We show how all three estimators can be extended to relax this assumption in the main equation, maintaining only the strict exogeneity of the regressors in the selection equation. We apply the extensions of the estimators to our particular problem, and compare the emerging estimates.

Another problem which frequently occurs with panel data is measurement error in some of the explanatory variables. With most panel surveys, the construction of work history variables needs to be based on retrospective information, which is likely to suffer from measurement error. If the affected variables enter the equation of interest in a non-linear manner, IV estimation does not solve the problem. We show how to address this problem within the methods discussed.

The data for our empirical application is drawn from the German Socio-Economic Panel (GSOEP). The dataset used for estimation is based on the first 12 waves of the panel.

The paper is organised as follows. In the next section we describe briefly the three estimators and their underlying assumptions. Section 3 compares the estimators. Section 4 discusses problems of implementation, and describes extensions to the case where strict exogeneity of some of the model regressors in the main equation is violated. Section 5 describes the data and the model we estimate. Section 6 presents the results, and Section 7 concludes.

2. The Model and Estimators

2.1 The model

The model we consider in the following consists of a binary selection rule, which depends on a linear index, and an unobserved (time constant) additive individual effect, which may be correlated with the model regressors. The selection rule assigns individuals in the overall sample population to two different regimes. For one regime, a linear regression equation is defined, which again has an additive unobserved individual component, correlated with the model regressors. The slope parameters of this equation are the parameters of interest.

This model can be written as:

$$w_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (2.1)$$

$$d_{it}^* = z_{it}\gamma - \eta_i - u_{it}; \quad d_{it} = 1[d_{it}^* \geq 0], \quad (2.2)$$

where $1[\cdot]$ is an indicator function, which is equal to one if its argument is true, and zero otherwise. Furthermore, β and γ are unknown parameter vectors, and x_{it}, z_{it} are vectors of explanatory variables with possibly common elements³, including both time variant and time invariant variables, and time effects. The α_i and η_i are unobservable and time invariant individual specific effects, which are possibly correlated with x_{it} and z_{it} . The ε_{it} and u_{it} are unobserved disturbances. The variable w_{it} is only observable if $d_{it} = 1$. The parameter vector we seek to estimate is β .

We assume that panel data is available. Equation (2.1) could be estimated in levels by pooled ordinary least squares (OLS). This leads to consistent estimates of β under the following condition:

³ For some estimators exclusion restrictions are not required because distributional assumptions (like normality of the error terms) identify the model. We assume throughout that there are exclusion restrictions in (2.1).

$$E(\alpha_i + \varepsilon_{it} | x_{it}, d_{it} = 1) = E(\alpha_i | x_{it}, d_{it} = 1) + E(\varepsilon_{it} | x_{it}, d_{it} = 1) = 0, \quad \forall t. \quad (2.3)$$

Accordingly, OLS estimates on the selected subsample are inconsistent if selection is non-random, and/or if correlated individual heterogeneity is present. In both cases, the conditional expectation in (2.3) is unequal to zero.

One way to eliminate the fixed effects α_i is to use some type of difference estimator. Given identification⁴, the consistency condition for an estimator using differences across time instead of level equations is given by the following expression:⁵

$$E(\varepsilon_{it} - \varepsilon_{is} | x_{it}, x_{is}, d_{it} = d_{is} = 1) = 0, \quad s \neq t, \quad (2.4)$$

where s and t are time periods.

Since condition (2.4) puts no restrictions on how the selection mechanism or the regressors relate to α_i , differencing equation (2.1) across time not only eliminates the problem of correlated individual heterogeneity but also any potential selection problem which operates through α_i .

If conditions (2.3) or (2.4) are satisfied, the OLS estimator or the difference estimator respectively lead to consistent estimates. No specification of the selection process is necessary. If conditions (2.3) and (2.4) are violated, consistent estimation requires to model the selection process. The estimators we describe in the next section take the consistency requirements (2.3) or (2.4) as a starting point. The idea of the estimator by Wooldridge (1995) is to derive an expression for the expected value in (2.3), and to add it as an additional regressor to the equation of interest. The estimator by Rochina-Barrachina (1999) derives an expression for the expected value in (2.4), which is then added as an additional regressor to the differenced equation. The estimator by Kyriazidou (1997)

⁴ For identification we require the matrix $E[(x_t - x_s)'(x_t - x_s) | x_t, x_s, d_t = d_s = 1]$ to be finite and non-singular.

⁵ If $s = t - 1$, the data is transformed by applying first differencing over time. Other transformations include mean deviation operators.

matches pairs of observations for a given individual for whom the conditional expectation in (2.4) is equal to zero.

2.2 Estimation in levels: Wooldridge's estimator

The estimation method developed by Wooldridge (1995) relies on level equations. The basic idea is to parameterise the conditional expectations in (2.3) and to add these expressions as additional regressors to the main equation. The method is semiparametric with respect to the main equation, in the sense that it does not require joint normality of the errors in both equations. Similar to Heckman's (1979) two-stage estimator, only marginal normality of the errors in the selection equation and a linear conditional mean assumption of the errors in the main equation is required. The time dimension allows controlling for individual effects in addition, which requires further assumptions for the conditional means of the individual effects in both equations. Wooldridge (1995) imposes two assumptions on the selection equation (**W1** and **W2** below), and two assumptions about the relationship between $\alpha_i, \varepsilon_{it}$ and the resulting error term in the selection equation (**W3** and **W4**).

- **W1:** *The regression function of η_i on z_i is linear.*

Following Chamberlain (1984), Wooldridge (1995) specifies the conditional mean of the individual effects in the selection equation as a linear projection on the leads and lags of the observable variables: $\eta_i = z_{i1}\delta_1 + \dots + z_{iT}\delta_T + c_i$, where c_i is a random component.

- **W2:** *The errors in the selection equation, $v_{it} = u_{it} + c_i$, are independent of \tilde{z}_i and normal $(0, \sigma_t^2)$, where $\tilde{z}_i = (x_i, z_i)'$ with $x_i = (x_{i1}, \dots, x_{iT})$ and $z_i = (z_{i1}, \dots, z_{iT})$.*⁶
- **W3:** *The regression function of α_i on x_i and v_{it} is linear.*⁷

⁶ v_{it} is heteroskedastic over time whenever u_{it} is.

⁷ An alternative assumption is (see Mundlack (1978), Nijman and Veerbeek (1992), and Zabel (1992)) that α_i depends only on the time average of x_{it} .

Accordingly, $E(\alpha_i | \tilde{z}_i, v_{it}) = x_{i1}\psi_1 + \dots + x_{iT}\psi_T + \phi_t v_{it}$.⁸ The conditional distribution of α_i on x_i, v_{it} is linear, but otherwise unrestricted. We do not observe v_{it} , however, but only the binary selection indicator d_{it} . Therefore, $E(\alpha_i | \tilde{z}_i, v_{it})$ has to be replaced by the expectation of α_i given $(\tilde{z}_i, d_{it} = 1)$, which is obtained by integrating $E(\alpha_i | \tilde{z}_i, v_{it}) = x_{i1}\psi_1 + \dots + x_{iT}\psi_T + \phi_t v_{it}$ over $v_{it} \leq z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{iT}$.⁹ This yields $E(\alpha_i | \tilde{z}_i, d_{it} = 1) = x_{i1}\psi_1 + \dots + x_{iT}\psi_T + \phi_t E[v_{it} | \tilde{z}_i, d_{it} = 1]$.

• **W4:** ε_{it} is mean independent of \tilde{z}_i conditional on v_{it} and its conditional mean is linear on v_{it} .

Accordingly, $E(\varepsilon_{it} | \tilde{z}_i, v_{it}) = E(\varepsilon_{it} | v_{it}) = \rho_t v_{it}$. The first equality states that ε_{it} is mean independent of \tilde{z}_i conditional on v_{it} , and the second equality that $E(\varepsilon_{it} | v_{it})$ is linear. No restrictions are imposed on the temporal dependence of ε_{it} , or on $\text{Corr}(\varepsilon_{it}, v_{is})$, for $s \neq t$. Again, as we do not observe v_{it} but the binary selection indicator d_{it} , we must find the expectation of ε_{it} given $(\tilde{z}_i, d_{it} = 1)$. This is obtained by integrating $E(\varepsilon_{it} | \tilde{z}_i, v_{it}) = \rho_t v_{it}$ over $v_{it} \leq z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{iT}$, resulting in $E(\varepsilon_{it} | \tilde{z}_i, d_{it} = 1) = \rho_t E[v_{it} | \tilde{z}_i, d_{it} = 1]$.

Under assumptions **W1–W4**, Wooldridge (1995) derives an explicit expression for

⁸ The key point for identifying the vector β is that, under v_{it} being independent of \tilde{z}_i , and the conditional expectation $E(\alpha_i | \tilde{z}_i, v_{it})$ being linear, the coefficients on the x_{it} , $r = 1, \dots, T$, are the same regardless of which v_{it} is in the conditioning set. This is crucial to the approach, and follows from the law of iterated expectations. For any t ,

$$\begin{aligned} E(\alpha_i | \tilde{z}_i) &= x_{i1}\psi_1 + \dots + x_{iT}\psi_T + \phi_t E[v_{it} | \tilde{z}_i] \\ &= x_{i1}\psi_1 + \dots + x_{iT}\psi_T \\ &= x_{i1}\psi_1 + \dots + x_{iT}\psi_T. \end{aligned}$$

The second equality follows because $E[v_{it} | \tilde{z}_i] = 0$ under **W2**, and the third follows from the coefficients in the linear projection of α_i onto x_i being necessarily time-invariant.

⁹ $z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{iT}$ is the reduced form index for the selection equation in (2.2), once the time-constant unobserved effect η_i is specified as in **W1**.

$$E(\alpha_i + \varepsilon_{it} | \tilde{z}_i, d_{it} = 1) = E(\alpha_i | \tilde{z}_i, d_{it} = 1) + E(\varepsilon_{it} | \tilde{z}_i, d_{it} = 1) = x_{i1}\psi_1 + \dots + x_{iT}\psi_T + (\phi_t + \rho_t) \cdot E[v_{it} | \tilde{z}_i, d_{it} = 1] \quad (2.3')$$

which results in the following model:

$$w_{it} = x_{i1}\psi_1 + \dots + x_{iT}\psi_T + x_{it}\beta + \lambda(H_{it}/\sigma_t) + e_{it}, \quad (2.5)$$

where $\sigma_t = \phi_t + \rho_t$, $H_{it} = z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{tT}$ is the reduced form index in the selection equation for period t , and $\lambda(H_{it}/\sigma_t) = E[v_{it} | \tilde{z}_i, d_{it} = 1]$.

Notice that, since $d_{it} = 1$ for $r \neq t$ is not included in the conditioning sets of $E(\alpha_i | \tilde{z}_i, d_{it} = 1)$ and $E(\varepsilon_{it} | \tilde{z}_i, d_{it} = 1)$, the selection term $E[v_{it} | \tilde{z}_i, d_{it} = 1]$ is not strictly exogenous in (2.5). The condition, which holds for the new error term in (2.5), is $E(e_{it} | \tilde{z}_i, d_{it} = 1) = E(e_{it} | \tilde{z}_i, v_{it} \leq H_{it}) = 0$. We call this a ‘‘contemporaneous exogeneity’’ of the selection term $E[v_{it} | \tilde{z}_i, d_{it} = 1]$ with respect to e_{it} in (2.5).

To obtain estimates for $\lambda(\cdot)$, a probit on $H_{it} = z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{tT}$ is estimated for each t in the first step. In the second step, equation (2.5) is estimated either by minimum distance or pooled OLS regression. Under the assumptions **W1-W4**, the estimator for β is consistent. Since dependence between the unobservables in the selection equation, v_{it} , and the unobservables in the main equation, $(\varepsilon_{it}, \alpha_i)$, is allowed for, selection may depend not only on the error ε_{it} , but also on the unobserved individual effect α_i . For time invariant variables or variables that vary systematically over time, β is not separable from ψ . For time varying variables we can identify β given that the coefficients ψ_1, \dots, ψ_T are constant for different time periods (assumption **W3**).

2.3 Estimation in differences I: Kyriazidou's estimator

The estimator developed by Kyriazidou (1997) relies on pairwise differences over time applied to model (2.1) for individuals satisfying $d_{it} = d_{is} = 1, s \neq t$. The idea of the estimator is as follows. Re-consider first the expression in (2.4):

$$\begin{aligned} E(\varepsilon_{it} - \varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) = \\ E(\varepsilon_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) - E(\varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) \equiv \\ \lambda_{its} - \lambda_{ist} \end{aligned} \quad (2.4')$$

where $\tilde{z}_{it} = (x_{it}, z_{it})'$, $\tilde{z}_{is} = (x_{is}, z_{is})'$, and for each time period the selection terms are

$$\begin{aligned} \lambda_{its} &= E(\varepsilon_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, u_{it} \leq z_{it}\gamma - \eta_i, u_{is} \leq z_{is}\gamma - \eta_i) \\ &= \Lambda(z_{it}\gamma - \eta_i, z_{is}\gamma - \eta_i; F(\varepsilon_{it}, u_{it}, u_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i)) \\ \lambda_{ist} &= E(\varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, u_{is} \leq z_{is}\gamma - \eta_i, u_{it} \leq z_{it}\gamma - \eta_i) \\ &= \Lambda(z_{is}\gamma - \eta_i, z_{it}\gamma - \eta_i; F(\varepsilon_{is}, u_{is}, u_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i)) \end{aligned}$$

where $\Lambda(\cdot)$ is an unknown function and $F(\cdot)$ is an unknown joint conditional distribution function of the errors. The additional variables in the conditioning set in (2.4'), compared to the conditioning set in expression (2.4), follow from the fact that the sample selection mechanism has to be specified in this model. The individual effects in both equations are allowed to depend on the explanatory variables in an arbitrary way, and are not subject to any distributional assumption. Different to Wooldridge (1995), the individual effects are now included in the conditioning set.

Under the assumption that for individuals for whom $z_{it}\gamma = z_{is}\gamma$ and $d_{it} = d_{is} = 1$, the sample selection effect is equal in t and s (that is, $\lambda_{its} = \lambda_{ist}$ in (2.4')). Hence, differencing between periods s and t will entirely remove the sample selection problem and, at the same time, the time constant individual heterogeneity component.

In general however there is no reason to expect that $\lambda_{its} = \lambda_{ist}$ holds even for individuals satisfying the conditions above. In particular, the selection terms depend not only on the conditioning vector $(\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i)$, but also on the joint conditional distribution of the error terms for the two time periods, which may differ across individuals, as well as over time for the same individual. To ensure that $\lambda_{its} = \lambda_{ist}$ holds, Kyriazidou (1997) imposes a “conditional exchangeability” assumption. The resulting estimator is semiparametric with respect to both the error distribution and the distribution of the fixed effects.

To implement this estimator, Kyriazidou (1997) imposes the following conditions:

• **K1:** $(\varepsilon_{it}, \varepsilon_{is}, u_{it}, u_{is})$ and $(\varepsilon_{is}, \varepsilon_{it}, u_{is}, u_{it})$ are identically distributed conditional on $\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i$. That is, $F(\varepsilon_{it}, \varepsilon_{is}, u_{it}, u_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i) = F(\varepsilon_{is}, \varepsilon_{it}, u_{is}, u_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i)$.

This “conditional exchangeability” assumption implies that the idiosyncratic errors are homoscedastic over time for a given individual. Under this assumption, any time effects are absorbed into the conditional mean.

• **K2:** An appropriate smoothness condition¹⁰ is imposed on the selection correction function $\Lambda(\cdot)$.

This smoothness condition ensures that once **K1** holds, $z_{it}\gamma = z_{is}\gamma$ implies $\lambda_{its} = \lambda_{ist}$.

Under assumptions **K1-K2** and provided identification is met,¹¹ the OLS estimator applied to

$$w_{it} - w_{is} = (x_{it} - x_{is})\beta + e_{its}, \quad (2.6)$$

¹⁰ Kyriazidou (1997) imposes a Lipschitz continuity property on the selection correction function $\Lambda(\cdot)$.

¹¹ In this model identification of β requires $E[(x_t - x_s)'(x_t - x_s)d_t d_s | (z_t - z_s)\gamma = 0]$ to be finite and non-singular. Given that we require support of $(z_t - z_s)\gamma$ at zero, nonsingularity requires an exclusion restriction on the set of regressors, namely that at least one of the variables z_{it} is not contained in x_{it} .

for individuals satisfying $d_{it} = d_{is} = 1, s \neq t$ and $z_{it}\gamma = z_{is}\gamma$, is consistent. The resulting error $e_{its} \equiv (\varepsilon_{it} - \varepsilon_{is}) - (\lambda_{its} - \lambda_{ist})$ has a conditional expectation that satisfies $E(e_{its} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) = 0$.

The estimator requires that there are individuals with $z_{it}\gamma = z_{is}\gamma$ with probability one, which is not the case if z_{it} contains a continuous variable. To implement the estimator, Kyriazidou (1997) constructs kernel weights, which are a declining function of the distance $|z_{it}\gamma - z_{is}\gamma|$, and estimates pairwise differenced equations by weighted OLS¹².

The procedure requires estimates of γ , which can be obtained either by smoothed conditional maximum score estimation (see, for instance, Charlier, Melenberg and van Soest (1997) and Kyriazidou (1997))¹³ or conditional logit (Chamberlain (1980)) estimation.

2.4 Estimation in Differences II: Rochina-Barrachina's estimator

This estimator is also based on pairwise differencing equation (2.1) for individuals satisfying $d_{it} = d_{is} = 1, s \neq t$. Different from Kyriazidou's (1997) estimator, Rochina-Barrachina's (1999) estimator relies on a parameterisation of the conditional expectation in (2.4). On the other hand, it does not impose the "conditional exchangeability" assumption.

To implement the estimator, the following assumptions are made:

- **RB1:** *The regression function of η_i on z_i is linear*¹⁴.
- **RB2:** *The errors in the selection equation, $v_{it} = u_{it} + c_i$, are normal $(0, \sigma_t^2)$.*

¹² The estimator is arbitrarily close to root n-consistency depending on the degree of smoothness one is willing to assume for the kernel function.

¹³ Estimating γ by the smoothed conditional maximum score estimator requires additional assumptions (see Manski (1987), Horowitz (1992), Kyriazidou (1994) and Charlier, Melenberg and van Soest (1997) for details).

In Rochina-Barrachina (1999), a non-parametric specification of the conditional mean of η_i is used, where $E(\eta_i | z_i)$ is left unrestricted.

• **RB3**: The errors $\left[(\varepsilon_{it} - \varepsilon_{is}), v_{it}, v_{is} \right]$ are trivariate normally distributed conditional on \tilde{z}_i .

The first two assumptions refer to the selection equation and are equivalent to assumptions **W1** and **W2** above. The third assumption imposes restrictions on the joint conditional distribution of the error terms in the two equations. The method is non-parametric with respect to the individual effects in the main equation and allows, under its semi-parametric version, for a non-parametric conditional mean of the individual effects in the selection equation on the leads and lags of the explanatory variables in that equation.

Under assumptions **RBI-RB3**, the resulting estimation equation is given by

$$w_{it} - w_{is} = (x_{it} - x_{is})\beta + {}_{ts}\lambda \left| \frac{H_{it}}{\sigma_t}, \frac{H_{is}}{\sigma_s}, \rho_{ts} \right| + {}_{st}\lambda \left(\frac{H_{is}}{\sigma_s}, \frac{H_{it}}{\sigma_t}, \rho_{ts} \right) + e_{its}, \quad (2.7)$$

where $H_{it} = z_{i1}\gamma_{\tau 1} + \dots + z_{iT}\gamma_{\tau T}$, $\tau = t, s$, are the resulting reduced form indices in the selection equation for periods t and s , and $\rho_{ts} = \rho_{(v_t/\sigma_t)(v_s/\sigma_s)}$ is the correlation coefficient between the errors in the selection equation. Furthermore,

${}_{ts}\lambda \left| \frac{H_{it}}{\sigma_t}, \frac{H_{is}}{\sigma_s}, \rho_{ts} \right| + {}_{st}\lambda \left(\frac{H_{is}}{\sigma_s}, \frac{H_{it}}{\sigma_t}, \rho_{ts} \right)$ is the conditional mean

$E(\varepsilon_{it} - \varepsilon_{is} | \tilde{z}_i, d_{it} = d_{is} = 1)$ derived from the three-dimensional normal distribution assumption in **RB3**.¹⁵ The new error term $e_{its} \equiv (\varepsilon_{it} - \varepsilon_{is}) - [{}_{ts}\lambda_{its} + {}_{st}\lambda_{ist}]$ has a conditional expectation $E(e_{its} | \tilde{z}_i, v_{it} \leq H_{it}, v_{is} \leq H_{is}) = 0$. To construct estimates of the $\lambda(\cdot)$ terms the reduced form coefficients (γ_t, γ_s) will be jointly determined with ρ_{ts} , using a bivariate probit for each combination of time periods. The second step is carried out by applying OLS to equation (2.7).

¹⁵ See Rochina-Barrachina (1999) for details.

3. Comparison of Estimators

Table 1 summarises the main features of the three estimators, and the assumptions they impose on the data. Wooldridge's (1995) method is the only one that relies on level equations. This makes it necessary to specify the functional form for the conditional mean of the individual effects in the main equation, α_i , with respect to the explanatory variables (to allow for individual correlated heterogeneity) and with respect to the random error term v_{it} (to allow for selection that depends on the unobserved effect α_i). In the other methods, α_i is differenced out, and selection may therefore depend on α_i in an *arbitrary* fashion.

With respect to the assumptions on the functional form of the sample selection effects, Kyriazidou (1997) treats them as unknown functions, which need not to be estimated. Wooldridge (1995) and Rochina-Barrachina (1999) parameterise these effects, which imposes three assumptions. First, a normality assumption for the random component of the unobservables in the selection equation ($v_{it} = c_i + u_{it}$). Secondly, to explicitly modelling the dependence of η_i on the explanatory variables. Thirdly, an assumption about the relationship between the errors in the main equation and the v_{it} in the selection equation. In Wooldridge (1995) joint normality of unobservables in both equations is not needed once a marginal normality assumption for the v_{it} and a linear projection specification for ε_{it} on v_{it} are imposed. In Rochina-Barrachina's (1999) estimator, joint normality is assumed, and linearity between ε_{it} and v_{it} results from the joint normality assumption.

Kyriazidou (1997) does not impose any parametric assumption about the distribution of the unobservables in the model. However, the conditional exchangeability assumption in Kyriazidou's (1997) estimator imposes restrictions on the time series properties of the model. This assumption is more demanding than joint conditional stationarity for the time-varying errors (see Kyriazidou (1997) for details). While in Wooldridge (1995) and Rochina-Barrachina (1999) not only the conditional means, but also the second moments of the error terms may incorporate time effects, Kyriazidou's (1997) estimator allows only for time effects in the conditional mean.

All these methods do not impose explicitly restrictions on the pattern of serial-correlation in the error processes. However, in Kyriazidou (1997) serial correlation is allowed as far as this does not invalidate the “conditional exchangeability” assumption. Wooldridge’s (1995) method imposes no restriction on the way the time-varying error in the main equation (ε_{it}) relates to the time-varying error in the selection equation (v_{is}), for $s \neq t$. Different to Wooldridge (1995), in Rochina-Barrachina’s (1999) estimator the joint normality assumption (**RB3** above) extends linearity to the correlation between ε_{it} and v_{is} for $s \neq t$, since it includes in the conditioning set not only d_{it} , but d_{it} , d_{is} .

The estimators differ in terms of sample requirements. In Wooldridge (1995) the parameters of interest are estimated from those observations that have $d_{it} = 1$. Rochina-Barrachina’s (1999) estimator uses individuals with $d_{it} = d_{is} = 1$. Kyriazidou (1997) uses those observations that have $d_{it} = d_{is} = 1$, and for which $z_{it}\gamma$ and $z_{is}\gamma$ are “close”. Asymptotically, the effective sample size is smaller for the latter method.

At the stage of implementation, problems may arise with Kyriazidou’s (1997) method if there are strong time effects in the selection equation. In this case, it may be difficult to find observations for which $z_{it}\gamma$ and $z_{is}\gamma$ are “close”. Furthermore, identification problems arise if for individuals for whom $z_{it}\gamma$ and $z_{is}\gamma$ are “close”, also x_{it} is “close” to x_{is} . In this case, a higher weight is given to observations with little time-variation in the explanatory variables in the main equation. A related problem arises if high matching weights are assigned to observations whose x variables change in a systematic manner. In this case it is not possible to separately identify the coefficients of these variables from coefficients on a time trend, or time dummies. These problems are likely to occur in many empirical applications, as we demonstrate below.

TABLE 1: COMPARISON OF ESTIMATORS

| Estimators | Estimation | Sample selection effects | Distributional assumptions | | | | Specification of conditional means | | |
|--------------------|------------|--------------------------|----------------------------|-------------------------------|--------------------------|--------------------------|--|--|--------------------------------|
| | | | α_i | η_i | ε_{it} | u_{it} | α_t | η_t | ε_{it} |
| Wooldridge | Levels | Parameterized | None | Normal random component c_i | None | Normal | LP ^a on x_i & $v_{it}=c_i+u_{it}$ | LP ^a on z_i | LP ^a on v_{it} |
| Kyriazidou | Time diff. | Unspecified | None | None | None but CE ^b | None but CE ^b | None | None | None |
| Rochina-Barrachina | Time diff. | Parameterized | None | Normal random component c_i | Normal | Normal | None | LP ^a on z_i /non-parametric | Linearity from joint normality |

| Estimators | Time series properties | | | | Sample requirements |
|--------------------|----------------------------|----------------|--------------------|---|---|
| | Time dummies or time trend | Time Heterosk. | Serial correlation | Corr(ε_{it}, u_{is}) $t \neq s$ | |
| Wooldridge | Yes | Yes | Yes | Unspecified | $d_{it} = 1$ |
| Kyriazidou | Yes | No | CE ^b | CE ^b | $d_{it} = d_{is} = 1,$ $z_{it}\gamma \cong z_{is}\gamma$ |
| Rochina-Barrachina | Yes | Yes | Yes | subject to joint normality | $d_{it} = d_{is} = 1$ |

^a LP denotes the linear projection operator.

^b Subject to the “conditional exchangeability” (CE) assumption according to which the vectors of errors $(\varepsilon_{it}, \varepsilon_{is}, u_{it}, u_{is})$ and $(\varepsilon_{is}, \varepsilon_{it}, u_{is}, u_{it})$ are identically distributed conditional on $\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i$.

4. Extensions

4.1 Estimation if regressors are non-strictly exogenous

All the estimators above assume strict exogeneity of the regressors. The variable x_{it} is strictly exogenous relative to ε_{it} if

$$E(\varepsilon_{it} | x_{it}) = 0, \quad t = 1, \dots, T. \quad (4.1)$$

A similar statement can be made about z_{it} with respect to u_{it} . If $E(\varepsilon_{it} | x_{it}) = 0$, we call this contemporaneous exogeneity.

In many empirical applications, the strict exogeneity condition (after controlling for both individual heterogeneity and sample selection) is likely to be violated. In the following, we describe how the above three estimators can be extended in this direction. We maintain the strict exogeneity assumption of regressors in the selection equation.

In Wooldridge (1995), the selection correction proposed has been derived under the assumption of strict exogeneity of the regressors conditional on the unobserved effect, that is, $E(\varepsilon_{it} | x_{it}, \alpha_i) = 0$. The strict exogeneity assumption is, for instance, needed for condition **W3** to be valid. To see this, suppose that the variables in the equation of interest are predetermined, and possibly correlated with the individual effects α_i . In this case, the set of valid conditioning variables for the linear projection of α_i on the regressors differs for different time periods – in period t the conditioning set is the vector $x_{it}^t \equiv (x_{it1}, \dots, x_{itT})$. If however the conditioning set changes over time, the coefficients for the leads and lags of the explanatory variables in the linear projection of α_i will likewise vary over time, thus invalidating **W3**. Hence, the condition for β to be separately identified from ψ (implying that $\psi_{t1} = \psi_1, \dots, \psi_{tT} = \psi_T, t = 1, \dots, T$) does not hold.

With pre-determined variables, identification of β requires the assumption that the variables in the main equation are not correlated with the individual effects α_i . Assumption **W3** is then substituted by

$$E(\alpha_i | \tilde{z}_i, d_{it} = 1) = q + \phi_t E(v_{it} | \tilde{z}_i, d_{it} = 1). \quad (4.2)$$

For many applications, this assumption is very restrictive.

One way to relax this assumption is to substitute the non-strictly exogenous time-varying correlated regressors by their predictions, and to apply Wooldridge's (1995) estimator. The construction of these predictions is not straightforward, however. For all time periods and for each non-strictly exogenous variable, T unique predictions are required. To identify β , assumption **W3** must hold. Accordingly, predictions for x_i for period t can not be constructed by using the subsample of individuals who participate during that period, where the instruments are both the sample selection term for that period (λ_{it}) and the leads and lags of the explanatory variables in the sample selection equation. This would produce multiple predictions for the same x_i in different time periods, thus invalidating **W3**. Also, we do not obtain unique predictions for x_i for all periods by including all the sample selection terms in the conditioning set, because the lambda terms are not strictly exogenous in the equation of interest (see discussion above). The way to obtain unique predictions is to predict each component of the vector x_i , using the entire sample of individuals in the participation equation, and all leads and lags of the explanatory variables in that equation as instruments.

The other two estimators rely on difference estimation. Hence pre-determined regressors in the level equation lead to endogenous regressors in the difference equation. In Kyriazidou's (1997) method, a straightforward way to allow for endogenous regressors is an IV type procedure¹⁶. Let z_i be the set of instrumental variables. Then the difference ($x_{it} - x_{is}$)

¹⁶ The IV version of Kyriazidou's (1997) estimator has been proved to be consistent in Charlier, Melenberg and Van Soest (1997).

fitted by z_i is $(\hat{x}_{it} - \hat{x}_{is}) = z_i' \sum_j z_j z_j' z_j (x_{jt} - x_{js})$, and the IV estimator b_{IV} has the form

$$b_{IV} = \sum_i (\hat{x}_{it} - \hat{x}_{is})' (x_{it} - x_{is}) d_{it} d_{is} \bar{\omega}_{ts} [(z_{it} - z_{is}) \hat{\gamma}]^{-1} \sum_i (\hat{x}_{it} - \hat{x}_{is})' (w_{it} - w_{is}) d_{it} d_{is} \bar{\omega}_{ts} [(z_{it} - z_{is}) \hat{\gamma}] \quad (4.3)$$

where $\bar{\omega}_{ts} [(z_{it} - z_{is}) \hat{\gamma}]$ is the kernel weight for individual i in pair (t, s) . This approach allows to maintain the same dimension of $(x_{it} - x_{is})$ in the estimated instrument set $(\hat{x}_{it} - \hat{x}_{is})$, which is computationally convenient. The pre-estimation of instruments does not affect the second-stage variance.

Given the non-parametric nature of the sample selection terms in this method, identification of the parameters of interest requires some component of z_{it} to be excluded from both the main equation and the instrument set. In practical applications, to find such variables can be hard.

The assumption of strictly exogenous regressors in the main equation for Rochina-Barrachina's (1999) estimator can be relaxed by applying a generalised method of moments estimator of the form

$$b_{GMM} = \sum_i \ddot{x}_{its} \ddot{z}'_{its} \Omega^{-1} \sum_i \ddot{z}_{its} \ddot{x}'_{its} \sum_i \ddot{x}_{its} \ddot{z}'_{its} \Omega^{-1} \sum_i \ddot{z}_{its} (w_{it} - w_{is}), \quad (4.4)$$

where $\ddot{x}_{its} \equiv [(x_{it} - x_{is}), \lambda_{its}, \lambda_{ist}]'$ and $\ddot{z}_{its} \equiv (z'_i, \lambda_{its}, \lambda_{ist})$. The matrix Ω is given by $\Omega = \sum_i \ddot{z}_{its} \ddot{z}'_{its} r_{its}^2$, where $r_{its} = (w_{it} - w_{is}) - (x_{it} - x_{is}) b^{IV} - [\lambda_{its}^{IV} + \lambda_{ist}^{IV}]$ are the estimated residuals. The z_i are defined as above, but now the instrument vector for a given pair (t, s) , \ddot{z}_{its} , also includes the corresponding sample selection terms λ_{its} and λ_{ist} . By

setting $\Omega = \sum_i \ddot{z}_{its} \ddot{z}'_{its}$ the GMM estimator becomes a simple IV estimator, and estimates can be used as initial estimates for the GMM estimator.

To summarise, if regressors in the main equation are non-strictly exogenous, the methods of Kyriazidou (1997) and Rochina-Barrachina (1999) may easily be extended to using IV or GMM type estimators. For Wooldridge's (1995) estimator, one solution of the problem is to use predicted regressors.

4.2 Measurement error

In typical panel surveys, the construction of work history variables, like tenure and experience, is based on retrospective information, which is likely to suffer from measurement error. An example is labour market experience, which is updated quite precisely during the course of the panel, but where the pre-sample information stems from retrospective data. The measurement error in this case is constant within individuals. If this variable enters the equation of interest in a linear way, differencing eliminates the measurement error. If this variable enters in a non-linear way (for instance, by including squared terms), differencing over time does not eliminate the measurement error, but it eliminates the problem associated to it.

To illustrate this, suppose that the variable x_{it} is measured with error, and we include its level and its square among the regressors in equation (2.1). Let the measured variable x^*_{it} be equal to the true variable x_{it} , plus an individual specific error term:

$$x^*_{it} = x_{it} + e_i, \tag{4.5}$$

where e_i is assumed to be uncorrelated with x_{it} . For Wooldridge's (1995) estimator, writing the true regression equation in (2.5) in terms of the observed variables leads to the following expression:

$$w_{it} = x_{i1}^* \psi_1 + \dots + x_{iT}^* \psi_T + x_{i1}^{*2} \Psi_1 + \dots + x_{iT}^{*2} \Psi_T + x_{it}^* \beta_1 + x_{it}^{*2} \beta_2 + \lambda(H_{it}/\sigma_t) + \left[e_{it} - (\psi_1 + \dots + \psi_T + \beta_1) e_i + (\Psi_1 + \dots + \Psi_T + \beta_2) e_i^2 - 2(\Psi_1 x_{i1}^* + \dots + \Psi_T x_{iT}^* + \beta_2 x_{it}^*) e_i \right] \quad (4.6)$$

where the new error term is now given by the expression in brackets.

A common solution to solve the measurement error problem is to use instrumental variable estimation. However, this estimation strategy does not longer lead to consistent estimates in a non-linear error in variables problem, because the error of measurement is no longer additively separable from the regressors (see expression (4.6)). Hence, it is impossible to find instruments which are correlated with the observed regressors, but uncorrelated with the new error term in (4.6).

An alternative solution is to use predicted regressors. In contrast to standard instrumental variables techniques, the use of predicted regressors, once the disturbances of the equation of interest have been purged for correlated heterogeneity and sample selection, allows to estimate the model under some conditions.

Let the true variable x_{it} be determined by a vector of instruments Z_i ,

$$x_{it} = Z_i \delta_t + s_{it}. \quad (4.7)$$

Assume that δ_t is known since it is identified from

$$x_{it}^* = Z_i \delta_t + s_{it} + e_i. \quad (4.8)$$

For Wooldridge's (1995) estimator, substitution of (4.7) into equation (2.5) yields the following expression

$$w_{it} = (Z_i \delta_1) \psi_1 + \dots + (Z_i \delta_T) \psi_T + (Z_i \delta_1)^2 \Psi_1 + \dots + (Z_i \delta_T)^2 \Psi_T + (Z_i \delta_1) \beta_1 + (Z_i \delta_1)^2 \beta_2 + \lambda(H_{it}/\sigma_t) + \left[e_{it} + (s_{i1} \psi_1 + \dots + s_{iT} \psi_T + s_{it} \beta_1) + (s_{i1}^2 \Psi_1 + \dots + s_{iT}^2 \Psi_T + s_{it}^2 \beta_2) + 2((Z_i \delta_1) s_{i1} \Psi_1 + \dots + (Z_i \delta_T) s_{iT} \Psi_T + (Z_i \delta_1) s_{it} \beta_2) \right],$$

where the term in brackets is the new error term. The assumption that s_{it} is independent of Z_i is crucial for consistent estimation, and necessary because of the non-linear specification. Independence guaranties not only that the first conditional moment of s_{it} is equal to zero, but also that the second conditional moment equals zero. Hence, we obtain an expression with linear and quadratic terms in $Z_i\delta_t$, for $t = 1, \dots, T$, and a new error term that is a function of the original error term, of linear and quadratic terms in s_{it} , and of cross products $s_{it}(Z_i\delta_t)$. To obtain consistent estimates, one needs to assume that $E(\text{new error term}|Z_i\delta_t) = 0$, implying that the Z_i are uncorrelated with the original error term in the equation of interest, and the s_{it} are independent of Z_i .

If estimating the model in differences (as in Kyriazidou (1997) or Rochina-Barrachina (1999)), and writing the true regression equation in (2.6) and (2.7) in terms of the observed variables in (4.5) we obtain:

$$\begin{aligned} w_{it} - w_{is} &= (x_{it}^* - x_{is}^*)\beta_1 + (x_{it}^{*2} - x_{is}^{*2})\beta_2 + E(\varepsilon_{it} - \varepsilon_{is}|\cdot) + [e_{its} - 2\beta_2(x_{it}^* - x_{is}^*)e_i] \\ &= (x_{it} - x_{is})\beta_1 + (x_{it}^{*2} - x_{is}^{*2})\beta_2 + E(\varepsilon_{it} - \varepsilon_{is}|\cdot) + [e_{its} - 2\beta_2(x_{it} - x_{is})e_i] \end{aligned} \quad (4.9)$$

where $E(\varepsilon_{it} - \varepsilon_{is}|\cdot)$ is equal to $E(\varepsilon_{it} - \varepsilon_{is}|\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1)$ for Kyriazidou (1997) and to $E(\varepsilon_{it} - \varepsilon_{is}|\tilde{z}_i, d_{it} = d_{is} = 1)$ for Rochina-Barrachina (1999). The new error is given by the term in brackets. Now the measurement error in $(x_{it}^{*2} - x_{is}^{*2})$ does not imply a measurement error problem for consistent estimation because e_i is uncorrelated with $(x_{it} - x_{is})$.¹⁷ Therefore, differencing eliminates the endogeneity problem due to measurement error, and the IV estimators in section 4.1 can be used¹⁸ to address the problem of non-strict exogenous regressors.

¹⁷ Since the error term in (4.9) includes $(x_{it} - x_{is})e_i$, and e_i is uncorrelated with $(x_{it} - x_{is})$, $E[(x_{it} - x_{is})e_i|(x_{it} - x_{is})] = E[(x_{it} - x_{is})e_i|(x_{it}^{*2} - x_{is}^{*2})] = 0$.

¹⁸ Although differencing within individuals does not eliminate the non-linear errors in variables, it does eliminate the problem. The quadratic terms, measured with error, are not longer endogenous to the new error

5. Empirical Model and Data

5.1 Estimation equation

We apply the estimators discussed in Section 2 to analyse wage equations of females, using data from a twelve-year panel. We define the wage equation and the participation equation as:¹⁹

$$w_{it} = x_{it}\beta + Exp_{it}\xi + Exp_{it}^2\zeta + \alpha_i + \varepsilon_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T, \quad (5.1)$$

$$d_{it}^* = z_{it}\gamma - \eta_i - u_{it}; \quad d_{it} = 1[d_{it}^* > 0], \quad (5.2)$$

where d_{it} is an indicator variable, being equal to one if the individual participates. The variable d_{it}^* is a latent index, measuring the propensity of the individual to participate in the labour market. Our parameter of interest is the effect of actual labour market experience (Exp) on wages. The vector x_{it} is a subset of z_{it} that contains education and time dummies. The vector z_{it} contains in addition age and its square, three variables measuring the number of children in three different age categories, an indicator variable for marital status, an indicator variable for the husband's labour market state, and other household income. We consider the participation equation as a reduced form specification, where labour market experience is reflected by the children indicators, age, and the other regressors. We assume that all regressors in the participation equation are strictly exogenous. The wage variable w_{it} in (5.1) is only observable if $d_{it} = 1$ in (5.2).

Within this model, there are a number of potential sources of bias for the effects of the experience variable. First, unobserved heterogeneity. Unobserved worker characteristics such as motivation and ability or effort may be correlated with actual experience: if high ability workers have a stronger labour market attachment than low ability workers, OLS on equation (5.1) results in upward biased coefficients. Second, sample selection bias.

term in the equation. The crucial conditions for this to happen are that the measurement error is time-constant, and uncorrelated with the underlying true variables.

Sample selection occurs if unobservable characteristics affecting the work decision are correlated with the unobservable characteristics affecting the process determining wages. If these unobservable characteristics are correlated with the observables, then failure to control for them will lead to incorrect inference regarding the impact of the observables on wages. Third, experience is likely to be non-strictly exogenous, even after controlling for heterogeneity and sample selection. Labour market experience in any period t is an accumulation of weighted past participation decisions: $Exp_{it} = \sum_{s=1}^{t-1} r_{is} d_{is}$, where r_{is} is the proportion of time individual i allocates in period s to the labour market²⁰. In turn, participation depends on wage offers received. Accordingly, any shock to wages in period t affects the level of labour market experience in the future, thus violating condition (4.1). Furthermore, given the above formulation, past shocks to wages affect current experience also by altering the weights r_{is} . A final problem is measurement error. As typical in survey data, the experience variable is constructed as the sum of pre-sample retrospective information, and experience accumulated in each year of the survey (see data section for details). Experience updates constructed within the 12 years of the survey should only be marginally affected by miss-measurement, but the pre-sample experience information is likely to suffer quite considerably from measurement error. This results in measurement error in the experience variable, which is constant over time for a given individual.

5.2 Data and sample retained for analysis

Our data is drawn from the first 12 waves of the German Socio-Economic Panel (GSOEP) for the years 1984-1995 (see Wagner et al. (1993) for details on the GSOEP). We extract a sample of females between 20 to 64 years old, who have finished their school education, and who have complete data during the sample period on the variables in Table 2 (with the exception of wages for females who do not participate in a given period). We exclude individuals who are self-employed in any of the 12 years. We define an individual as participating in the labour market if she reports to have worked for pay in the month

¹⁹ See Appendix I for a motivation of this specification.

preceding the interview. We compute wages by dividing reported gross earnings in the month before the interview by the number of hours worked for pay. We obtain a final sample of 1053 individuals, resulting in 12636 observations. We use both participants and non- participants for the estimation of the selection equation. For estimating the wage equations, we use all females that participate in at least two waves.

Summary statistics and a more detailed description of the variables are given in Table 2. The variable *Exp*, which reports the total labour market experience of the individual in the year before the interview, is computed in two stages: First, we use information from a biographical scheme, which collects information on various labour market states before entering the panel. This information is provided on a yearly basis, and participation is broken down into part-time and full-time participation. We sum these two labour market states up to generate our total experience variable at entry to the panel. In every succeeding year, this information is updated by using information from a calendar, which lists labour market activities in every month of the year preceding the interview. Again, we sum up part-time and full-time work. Accordingly, after entering the panel, our experience variable is updated on a monthly basis. Furthermore, it relates to the year before the wage information is observed. If wage contracts are re-negotiated at the beginning of each calendar year, this experience information should be the information on which the current contract is based. Participation is defined as being in the state of part-time or full-time employment at the interview time. Non-participation is defined as being in the state of non-employment or unemployment. On average, 54 percent of our sample population participates in the labour market. The average age in the whole sample is 42 years, with individuals in the working sample being slightly younger than in the non-working sample.

We do not restrict our sample to married females. From the 12636 observations, 10680 (84.52 percent) are married females, of whose 51 percent participate in the labour market. We observe a higher percentage of labour market participants (72 percent) among the non-married. Of the 1053 females in our sample, 780 are married in each of the 12 periods, 87

²⁰ The process generating experience can be expressed as: $Exp_{it} = Exp_{it-1} + r_{it-1}d_{it-1}$, where by direct substitution we get $Exp_{it} = \sum_{s=1}^{t-1} r_{is}d_{is}$.

are not married in any period, and 186 are married between 1 and 11 years of the sample periods.

Our children variables distinguish between the number of children aged between 1 to 3 years, the number of children aged between 3 and 6 years, and the number of children between 6 and 16 years old. As one should expect, for all three categories, numbers are higher among the non-participants.

To estimate our wage equation conditional on fixed effects, we need repeated wage observations for the same individual. Table 3 reports frequencies of observed wages, as well as the number of state changes between participation and non-participation. 23 percent of our sample individuals participates in none of the 12 years, and about 25 percent in each of the 12 years. More than half of the sample has at least one state change within our observation window and there are no individuals who change state more than 7 times over the 12 years period. In the longitudinal dimension, 767 women (corresponding to 6757 observations) worked for a wage at least in two years during the sample period. Once we drop observations of individuals who do declare participation, but not wages, our number reduces to 5861 observations. The data we use for estimating the wage equation uses all individuals who report wages in at least two periods.

TABLE 2: DESCRIPTION OF VARIABLES AND SAMPLE STATISTICS (12,636 observations)^a

| Variable | Description | Total Sample | Work=1 (6802 observations) | Work=1 dropping individuals with participation in one year only and observations with missing wages (5861) | Work=0(5834 observations) |
|--------------------------|---|---------------------|----------------------------|--|---------------------------|
| Work | dummy variable indicating participation of the female (work=1) or no participation (work=0) | 0.538 (0.498) | 1 (0) | 1 (0) | 0 (0) |
| Lnwage | log gross hourly real wages (1984 West German Marks) | 2.681 (0.435) | 2.681 (0.435) | 2.685 (0.432) | |
| Exp | years-equivalent worked for money after leaving education | 14.373 (9.782) | 17.661 (9.407) | 17.931 (9.331) | 10.541 (8.765) |
| Exp2 | experience squared and divided by 10 | 30.231 (36.606) | 40.040 (38.264) | 40.861 (38.122) | 18.794 (30.862) |
| Time | time (year-1900), we also use time dummies for estimation | 89.500 (3.452) | 89.477 (3.435) | 89.457 (3.437) | 89.526 (3.472) |
| Age | age of the female in years | 42.263 (9.953) | 41.259 (9.356) | 41.205 (9.381) | 43.434 (10.487) |
| Age2 | Age of the female squared and divided by 10 | 188.527 (84.624) | 178.988 (76.917) | 178.592 (76.952) | 199.650 (91.567) |
| Ed | Education of the female measured as years of schooling | 10.847 (1.958) | 11.057 (2.129) | 11.103 (2.128) | 10.602 (1.705) |
| Hhinc | Additional real income per month (in thousands) | 2.735 (1.778) | 2.439 (1.855) | 2.394 (1.897) | 3.080 (1.617) |
| M | Dummy variable with value 1 if female married and value 0 if not married | 0.845 (0.361) | 0.793 (0.404) | 0.787 (0.409) | 0.905 (0.293) |
| hwork^b | Dummy variable with value 1 if husband works and value 0 if does not work | 0.862 (0.345) | 0.877 (0.328) | 0.875 (0.331) | 0.846 (0.361) |
| cc1 | Number of children up to 3 years old in the household | 0.117 (0.399) | 0.064 (0.301) | 0.059 (0.287) | 0.179 (0.481) |
| cc2 | Number of children between 3 and 6 years old in the household | 0.173 (0.442) | 0.118 (0.364) | 0.110 (0.351) | 0.238 (0.511) |
| cc3 | Number of children older than 6 years in the household | 0.436 (0.739) | 0.393 (0.696) | 0.366 (0.675) | 0.485 (0.784) |

^aStandard errors in parenthesis.^bThe reported sample statistics for this variable are conditional on the female being married.

TABLE 3: STATE FREQUENCIES

| No. of Years | Participating Individuals | | Number of State Changes | | |
|--------------|---------------------------|---------|-------------------------|-----------|---------|
| | Frequency | Percent | Changes | Frequency | Percent |
| 0 | 241 | 22.89 | 0 | 502 | 47.67 |
| 1 | 45 | 4.27 | 1 | 273 | 25.93 |
| 2 | 29 | 2.75 | 2 | 131 | 12.44 |
| 3 | 40 | 3.80 | 3 | 84 | 7.98 |
| 4 | 53 | 5.03 | 4 | 47 | 4.46 |
| 5 | 47 | 4.46 | 5 | 10 | 0.95 |
| 6 | 37 | 3.51 | 6 | 3 | 0.28 |
| 7 | 49 | 4.65 | 7 | 3 | 0.28 |
| 8 | 49 | 4.65 | | | |
| 9 | 59 | 5.60 | | | |
| 10 | 61 | 5.79 | | | |
| 11 | 82 | 7.79 | | | |
| 12 | 261 | 24.79 | | | |
| | 1053 | 100 | | 1053 | 100 |

TABLE 4: NUMBER OF OBSERVATIONS WORK=1 VERSUS WORK=0

| Years | Ratios Work=1/0 in participation sample | number of Work=1 dropping individuals with participation in one year only and observations with missing wages |
|--------------|---|---|
| 84 | 565/488 | 482 |
| 85 | 579/474 | 500 |
| 86 | 572/481 | 512 |
| 87 | 561/492 | 493 |
| 88 | 551/502 | 479 |
| 89 | 563/490 | 488 |
| 90 | 576/477 | 480 |
| 91 | 592/461 | 496 |
| 92 | 578/475 | 503 |
| 93 | 576/477 | 487 |
| 94 | 554/499 | 482 |
| 95 | 535/518 | 459 |
| 84-95 | 6802/5834 | 5861 |

6. Estimation Results

We concentrate most of our discussion on the effect of labour market experience. We use experience and its square as regressors in the wage equation. To facilitate the comparison of results in the various model specifications, we compute the rate of return to work experience

$$\partial w / \partial EXP = \xi + 2\zeta EXP, \quad (6.1)$$

where we evaluate the expression in (6.1) at 14 years of work experience (the sample average).²¹ We report estimates in Table 5. The full set of results is given in Table II.1 in the appendix. Rates of return implied by the different methods and for increasing levels of work experience are presented in Table II.2.

Column (1) presents OLS estimates, where we allow for time effects, but no individual effects. The results suggest that, evaluated at 14 years of labour market experience, an additional year increases wages by 1.48 percent. If high ability individuals have a stronger labour market attachment than low ability individuals, then this estimate should be upward biased. Furthermore, sample selection should re-enforce this upward bias if unobservables determining participation are positively correlated with unobservables in the wage equation (either through the α_i or the ε_{it} terms).

In columns (2) and (3), we present estimators that difference out the fixed effects. Column (2) displays standard fixed-effects (within) estimates (FE), and column (3) difference estimates (DE), where all pair differences within time periods per individual are used.²² Both estimators allow for individual effects correlated with the explanatory variables. Thus, the upward bias induced by individual fixed effects and any sample selection bias acting through α_i should be eliminated. Interestingly,

²¹ Standard errors of this term are easily derived from the variances and covariances of the parameter estimates for ξ and ζ .

²² We estimate pooled OLS on 66 pairs corresponding to 25021 observations.

TABLE 5: Marginal Experience Effects, WAGE EQUATION^a

| | (1) OLS | (2) FE | (3) DE (OLS) | (4) DE (IV) | (5) DE (GMM) | (6) ^b W (MD) | (7) ^c W (MD) _(Exp) |
|---|---------------------|---------------------|---------------------|----------------------------------|----------------------------------|--------------------------------------|--|
| $\partial w / \partial EXP$ (14 years) | 0.0148* (0.0007) | 0.0223* (0.0056) | 0.0200* (0.0039) | 0.0340* (0.0054) | 0.0305* (0.0014) | 0.0148* (0.0077) | 0.0182* (0.0038) |
| Wald Test (Selection) | | | | | | $\chi^2_{42} =$ 17.22 (0.1412) | $\chi^2_{12} = 17.44$ (0.1336) |
| Wald Test (Fixed Effects) | | | | | | $\chi^2_2 = 6.03$ (0.049) | $\chi^2_2 = 5.66$ (0.062) |
| Hausman (Exo-geneity) | | | | $\chi^2_{14} = 92.84$ (0.000) | $\chi^2_{14} = 55.35$ (0.000) | | $\chi^2_{29} = 46.39$ (0.021) |

| | (8) ^d K | (9) ^d K (IV) | (10) ^c RB | (11) ^c RB (IV) | (12) ^e RB (GMM) |
|---|-----------------------------------|-------------------------------|---------------------------------------|---------------------------------------|--|
| $\partial w / \partial EXP$ (14 years) | 0.0409* (0.0105) | 0.0116 (0.0637) | 0.0129* (0.0054) | 0.0122* (0.0062) | 0.0097* (0.0017) |
| Hausman (Selection) | $\chi^2_2 =$ 6.6332 (0.036) | | | | |
| Wald Test (Selection) | | | $\chi^2_{132} =$ 292.60 (0.000) | $\chi^2_{132} =$ 311.04 (0.000) | $\chi^2_{132} =$ 3859.11 (0.000) |
| Wald-test (Exo-geneity) | | | | $\chi^2_{145} =$ 433.15 (0.000) | $\chi^2_{145} =$ 1241.19 (0.000) |

^a The numbers in parentheses below the coefficient estimates are standard errors. The numbers in parentheses below the test statistics are p-values.

^b Standard errors **corrected** for the first stage maximum likelihood probit estimates.

^c Standard errors **corrected** for the first stage maximum likelihood probit estimates and the use of predicted regressors.

^d Standard errors **corrected** for the prior in the time dummies coefficients.

^e Standard errors **corrected** for the first stage maximum likelihood bivariate probit estimates.

* Statistically different from zero at the five-percent significance level.

our estimates increase relative to the simple OLS estimations – point estimates for the fixed effect estimator and the difference estimator are 0.022 and 0.020 respectively.

An explanation for these increases in coefficients is measurement error. As we have shown above, differencing in a quadratic specification eliminates the effect of a time constant measurement error. If the downward bias of the experience coefficient in a level equation, induced by measurement error, is larger than the upward bias due to individual fixed effects, then the coefficient estimates of difference estimators should increase, compared to level estimation.

If past wage shocks affect current experience levels, then experience is not strictly exogenous in the wage level equation. Furthermore, it is endogenous in the difference equation. A common solution to this problem in standard difference estimators is to use instrumental variable techniques. Column (4) and (5) present results when applying IV and GMM techniques to our particular problem. These estimators are obtained by pooled IV and GMM on 66 pairs of combinations of time periods which we can form with a panel of 12 years²³. As instruments, we use all leads and lags of the variables in the sample selection equation. A Hausman-type test comparing the difference IV and GMM estimators with the differenced OLS estimator leads to rejecting exogeneity for the experience variables.

The estimates we obtain for the rate of return to work experience are slightly higher than those obtained with the difference estimators, with point estimates of 0.034 and 0.030 in the IV and GMM estimators respectively. This is consistent with experience being predetermined. If past positive shocks to wages increase the probability of past participation, then the coefficient on the experience variable should be downward biased in a simple difference equation under OLS estimation.

The (IV) difference estimates are consistent under the assumption that selection only works through the fixed effects. If however there is sample selection acting through ε , our instruments are invalid. In this case, the error term will incorporate the extra element:

²³ The IV estimates are used as the first step estimates to obtain the GMM estimates.

$$E(\varepsilon_{it} - \varepsilon_{is} | d_{it} = d_{is} = 1) \neq 0. \quad (6.2)$$

Clearly, a proper instrument set should be uncorrelated with the truncated conditional expectation in (6.2). In most applications, this is unlikely to be the case since the available instruments determine also the selection into the observed regime. For instance, in our case, every variable that affects the participation decision in previous periods should also affect the level of experience in the current period. In this case, $E\left\{\left[(\varepsilon_{it} - \varepsilon_{is}) + E(\varepsilon_{it} - \varepsilon_{is} | d_{it} = d_{is} = 1)\right] | z_i\right\} \neq 0$. Accordingly, a time variant selection process may invalidate instruments in a difference equation, if these instruments are correlated with the selection process.

We now turn to estimation results which take account of a selection process that operates both through ε and α , and we demonstrate how the problems of measurement error and pre-determinedness can be solved within this framework.

6.1 Wooldridge's estimator

Estimation results for Wooldridge's (1995) estimator are presented in columns (6) and (7). We have specified the conditional mean of the individual effects, following Mundlack (1978), as a linear projection on the within individual means of experience and its square. Results in column (6) are based on the assumption that experience is (strictly) exogenous. Results in column (7) allow for endogeneity by using predictions for the experience terms. This procedure takes care of both measurement error, and non-strict exogeneity.

Estimators in columns (6)-(7) are implemented as follows. After obtaining the selection terms by estimating probits for each wave, the wage equation in (2.5) is estimated for each time period. From these estimations, we obtain the unrestricted coefficients for the constant and education, 2 coefficients for the mean of experience and its square, the 2 coefficients of interest for experience and its square, and the coefficient for the selection correction term in a given period. In a second step, we use minimum distance to impose the cross-equation

restrictions. To obtain the predictions for the experience variable (results in column (7)), we predict the vector $(Exp_{i1}, \dots, Exp_{i12}, Exp_{i1}^2, \dots, Exp_{i12}^2)$ using the 1053 individuals in the sample selection equation, as well as all the leads and lags of the explanatory variables in that equation. The components of this vector of predictions are used to obtain the average predicted experience and its average predicted square.

The coefficient estimate for Wooldridge's (1995) estimator is 0.0148 (column 6), which is exactly equal to the OLS result. It is smaller than the fixed effects estimators in columns (2) and (3), which is to be expected if participation is selective and/or there is a measurement error problem in the level equation (which leads to a downward bias). To test for sample selection, we have performed a Wald test on the significance of the selection effects, where $H_0: \gamma = 0$. This test can be interpreted as a test of selection bias. However, the assumptions under the null hypothesis are stronger than what is required for simple fixed effects estimators, since $W3$ is maintained under H_0 ²⁴. The value for the test statistic is $\chi_{12}^2 = 17.22$, with a p-value of 0.1412. Thus, the null hypothesis can not be rejected. We also performed a Wald test for the joint significance of the ψ coefficients, where $H_0: \psi = 0$. The resulting value for the test statistic is larger than the critical value of the χ_2^2 , at the five-percent significance level, rejecting the null hypothesis, and indicating the presence of correlated fixed effects.

In column (7) we use predictions for the experience variables. This leads to an increase of the experience coefficient (from 0.014 in column (6) to 0.018 in column (7)). This indicates that there is endogeneity, induced by non-strict exogeneity of the experience variable, and/or measurement error. Hausman-type tests, comparing (6) and (7), lead to rejecting exogeneity both after controlling for correlated heterogeneity and sample selection. We perform a Wald tests for the estimates in column (7), testing the null hypotheses that $H_0: \gamma = 0$ and $H_0: \psi = 0$. Again, we cannot reject the null hypothesis $H_0: \gamma = 0$, but we reject the null hypothesis $H_0: \psi = 0$ at a 6.21 percent significance level.

6.2 Kyriazidou's estimator

To implement this estimator, we estimate in a first step a conditional logit fixed effects model (see Chamberlain, 1980). The results are displayed in column (4) of Table III.1 in Appendix III. These first step estimates are then used to calculate weights for the pairs of observations in the difference estimator. To construct the weights we use a normal density function for the kernel. We follow the plug-in procedure described by Horowitz (1992) to obtain the optimal kernel bandwidth.²⁵ Finally, we perform minimum distance to obtain the parameter estimates. The minimum distance estimator is the weighted average of the estimators for each pair, with weights given by the inverse of the corresponding covariance matrix estimate²⁶.

As discussed above, the estimator relies on a conditional exchangeability assumption that restricts the error terms to be homoscedastic over time. This assumption seems quite restrictive, in particular when estimating wage equations. There is strong evidence that the variance of the wage distribution has increased considerably over the last two decades. The assumption that the error terms in the selection equation are stationary over time is testable. Table III.1 displays results of the selection equation under the assumption of equal variances over time (column (2)), and estimates that relax this assumption (column (3)). A χ^2 test can be used to test for the joint conditional exchangeability assumption. The increment in the distance statistic²⁷ is 146.8201 with a p-value of 0.0002, which clearly leads to rejecting the null hypothesis (the test statistic is χ_{93}^2 distributed)²⁸. Therefore, the joint conditional exchangeability assumption is rejected for our application.

²⁴ See Wooldridge (1995) for details on this point.

²⁵ With this procedure, some initial value for the bandwidth is chosen. Then the parameter estimates, the estimate of the asymptotic bias and the estimate of the covariance matrix are computed. These estimates are used to compute the mean square error minimising bandwidths. We do a search among initial bandwidths, stopping the process when the chosen initial value of the bandwidth is close enough to the optimal one. As we estimate 66 panel wave pairs $t \neq s$, 66 optimal bandwidth are estimated.

²⁶ In principle, to estimate the optimal weighting matrix for the minimum distance will require estimates for the covariance matrix of the estimators for the different pairs of time periods. However, Charlier, Melenberg and Van Soest (1997) proof that these covariances converge to zero due to the fact that the bandwidth tends to zero as the sample size increases. As a consequence the optimal weighting matrix simplifies to a block diagonal matrix where each block corresponds to the inverse of the covariance matrix for a given panel wave pair.

²⁷ Testing for additional restrictions in minimum distance estimators can be found in Chamberlain (1984).

²⁸ The degree of freedom is 104 (the number of parameter estimates in the minimum distance for column (3)) minus 11 (the number of additional restrictions imposed in the minimum distance estimator of column (2)).

When applying this method to our data, we face a further problem: Asymptotically, the method uses only observations for which the index from the sample selection rule is the same in the two time periods. In our application, there are strong time effects in the selection equation. Furthermore, changes in the variable experience are strongly related to changes in our identifying instruments, like, for instance, the number of children. Any systematic increase in experience between two periods can not be distinguished from the time trend; any non-systematic change coincides with a change of variables in the selection equation. By the nature of the estimator, however, the latter pairs of observations obtain a small kernel weight, and they therefore contribute very little to identifying the experience effects. Hence, without further assumptions, we can not identify the experience effects. Similar identification problems are likely to occur in any application where unsystematic changes in the variable of interest coincide with differences in the index function used for constructing the weights. For our particular application, a possible solution to this problem is to use information on aggregate wage growth from other sources. To illustrate the estimator, we use here time effects obtained from simple difference estimators.

Estimation results are displayed in columns (8) and (9). In both specifications, we use time effects obtained from the difference estimator in column (3). Column (8) displays results of simple weighted OLS estimation of equation (6). The IV estimates presented in column (9) are obtained by following the procedure described for Kyriazidou's (1997) method in section 4.1 above.

As we already pointed out, given the non-parametric nature of the sample selection terms in this method, identification of the IV estimator requires at least one time-varying variable in the selection equation, which is to be excluded not only from the main equation, but also from the instrument set for experience. Such exclusions are difficult to justify in most circumstances. In our particular case, the experience variable measures the total labour market experience of the individual in the year before the interview. Since it is the weighted sum of past participation decisions, it should be explained by variables that influence past participation, like lags of the husband's income, and lagged children variables. Participation in the current period is affected by current variables (like

children(t), hhinc(t), etc.). This should identify the model. We need one exclusion restriction, and we exclude current other household income hhinc(t) from the instrument set for experience.

The estimator in (8) does not correct for possible endogeneity of the experience variable. The coefficient for the experience effect indicates that a year of labour market experience increases wages by 4.1 percent. This estimate is very large. The estimator in (9) corrects for non-strict exogeneity of the experience variable in the level equation. Instrumenting reduces the experience effect to 1.2 percent, but the effect is not statistically significant (which may be due to the smaller effective sample size used for this estimator). Because of the problems discussed above, we do not wish to overemphasise these estimates. Also, the estimates are obviously sensitive to the choice of pre-estimated time effects.

A Hausman-type test comparing the parameter estimates in column (8) with the difference estimator in column (3) indicates that the null hypothesis of no selectivity bias is rejected.

6.3 Rochina-Barrachina's estimator

Columns (10)–(12) present estimates, using the method by Rochina-Barrachina (1999). Column (10) displays results of simple OLS estimation of equation (2.7). IV-GMM estimates are presented in columns (11) and (12). For estimation, we use each combination of panel waves (t,s), resulting in a total of 66 pairs. To combine these estimates, we use minimum distance.²⁹ We obtain coefficients for 11 time dummies, the coefficients on experience and its square, and estimates of $66 \times 2 = 132$ coefficients for the correction terms for all the pairs.³⁰ The standard errors we present in table 5 are corrected for the first step bivariate probit estimates. The variables used as instruments are the leads and lags of the variables included in the sample selection equation, and the corresponding two sample selection terms of each pair of time periods.

²⁹ The optimal weighting matrix is obtained from an estimate for the covariance matrix of the estimators for the different time periods.

³⁰ The estimates can be obtained upon request.

The first step estimator of the parameters $\gamma_t, \gamma_s, \rho_{ts}$, which are used for constructing the correction terms, are obtained by estimating 66 bivariate probits. The parameters we estimate in each bivariate probit are the reduced form parameters of the corresponding indices of the selection rules for the two time periods. We also get an estimate of the correlation coefficient between the errors in the two time periods. The mean value for ρ_{ts} is 0.7862 (se=0.1299) with a minimum at 0.4845 and a maximum at 0.9658. Consequently, we reject on average $H_0: \rho_{ts} = 0$. Correlation appears because of the c_i component in the error term and/or because of serially correlated idiosyncratic errors.

We test whether the 66*2 correction terms are jointly significant, using Wald tests. The resulting values for the test statistics for the estimators in Columns (10) to (12) are clearly larger than the critical values of the χ^2_{132} at any conventional significance level. Hausman-type tests comparing the IV and the GMM estimators with the OLS estimator in Column (10) lead to rejecting exogeneity both after controlling for correlated heterogeneity and sample selection.

The estimated parameters are slightly lower than the OLS estimates, and do not differ very much between specifications. They indicate that, evaluated at 14 years of labour market experience, an additional year increases wages by about 1 percentage point. Compared to Wooldridge's (1995) estimator, estimates are slightly smaller, which may be due to different parametric assumptions imposed by the two estimators. Furthermore, estimates are remarkably similar across specifications. One reason for this similarity is that with Rochina-Barrachina's estimator, instrumenting corrects only for the non-strict exogeneity problem. With Wooldridge's (1995) estimator, the use of predicted regressors corrects also for the measurement error bias.

Interesting is also a comparison of wage growth due to aggregate time effects. In the last row of Table II.1, we display average wage growth for the 12 years period due to common time effects. The numbers indicate that the different methods result in different numbers. For instance, Rochina-Barrachina's estimator in column (10) assigns about 8 percent more wage growth over the 12 years period to time effects than the simple OLS estimator (column 3). An explanation for these differences is that Rochina-Barrachina's (1999)

method controls for time-varying sample selection (as does Wooldridge's estimator). As pointed out by Moffitt (1984), wages may trend not only because of aggregate wage growth (proxied by the time dummies), but also because of changes in the sample selection over time. If sample selection decreases over time, and if we do not control for selection, the time dummies will pick up this trend, leading to decreasing time effects in standard fixed effects and difference estimators, like the ones displayed in columns (2) to (5). This leads to downward biased time dummies. With Rochina-Barrachina's (1999) estimator, the time dummies will presumably pick up just the secular productivity growth, since it controls for the decline in sample selection over time.

With this method, the sample selection term is given by a parameterisation of the conditional mean $E(\varepsilon_{it} - \varepsilon_{is} | \tilde{z}_i, d_{it} = d_{is} = 1)$, $s < t$. We obtain for most individuals negative predictions for these expectations. To investigate whether sample selection does indeed decrease over time, we write the estimated values of these conditional means as a function of 11 time dummies in differences (after controlling for the increments in experience and its square). Using minimum distance estimation, we obtain negative and significant coefficients for the time dummies, which increase in absolute value over time. This indicates that sample selection does in fact decline over time.³¹

³¹ This result is in line with the estimates obtained for the participation equation in Appendix III. Here, the estimates for the time dummies show that female labour force participation increases over the length of the panel. Hence, as participation probabilities increase, sample selection may be reduced.

7. Conclusions

In many empirical applications, the equation of interest is defined for a non-random sample of the overall population. Furthermore, at the same time the outcome equation contains an unobserved individual specific component which is correlated with the model regressors. In this paper we discuss three estimators which may be applied if both problems occur simultaneously: The estimators of Wooldridge (1995), Kyriazidou (1997), and Rochina-Barrachina (1999). We investigate and compare the conditions under which they produce consistent estimates. We show how these estimators can be extended to take account of non-strict exogeneity and/or time constant non-linear errors in variables. We illustrate that, if regressors in the main equation suffer from these problems, the methods of Kyriazidou (1997) and Rochina-Barrachina (1999) can be straightforwardly extended to using IV or GMM type estimators. For Wooldridge's (1995) estimator, one solution of the problem is to use predicted regressors.

Not many applications exist for sample selection estimators in panel data models. To learn about the performance of the methods in a practical application, we apply the estimators and their extensions to a typical problem in labour economics: The estimation of wage equations for female workers. The parameter we seek to identify is the effect of actual labour market experience on wages. The problems that arise in this application are non-random selection, and unobserved individual specific heterogeneity which is correlated with the regressors. In addition, actual experience is predetermined, and the experience measure is likely to suffer from measurement error.

A flexible and attractive estimator is that by Kyriazidou (1997). It turns out however that, for our particular application, this estimator is difficult to apply. The estimator is very flexible in that it avoids specifying the sample selection terms, and it requires no parametric assumptions about the unobservables in the model. But it imposes a conditional exchangeability assumption, which is rejected by the data in our particular application. Furthermore, in the case where any non-systematic variation in the variable of interest (experience in our case) coincides with changes in the selection index, this estimator runs into identification problems (between time effects and experience in our case), that can only

be solved by using additional information. We use pre-estimated time dummies from simple difference estimators. To implement the IV estimator (which is producing consistent estimates if experience is pre-determined and/or contemporaneously endogenous), we need a further identification assumption. The estimate we obtain for the effect of labour market experience for the simple Kyriazidou estimator is quite large: Evaluated at 14 years of labour market experience, an additional year increases wages by about 4 percentage points. The estimates are sensitive to the pre-estimated time effects. The IV estimates are smaller, but not precisely estimated.

The results we obtain using Wooldridge's and Rochina-Barrachina's estimators indicate that there are correlated fixed effects, and non-random sample selection. With Wooldridge's (1995) estimator, the null hypothesis of no correlated fixed effects is rejected for all specifications. Conditional on fixed effects, the null hypothesis of no sample selection can not be rejected with Wooldridge's (1995) estimator, but it is clearly rejected with Rochina-Barrachina's (1999) estimator.³² Using Wooldridge's (1995) estimator, we reject specifications, which do not allow for predetermined regressors (and contemporaneous endogeneity). Rochina-Barrachina's (1999) method rejects strict exogeneity of the experience variable, conditional on taking care of the measurement error problem by time differencing. Accordingly, the use of sample selection models which take care of correlated fixed effects seems to be justified. Furthermore, the extensions we suggest in this paper seem to be important for our particular application.

The most general estimator using Wooldridge's (1995) method implies an increase in wages by 1.8 percent for one year of labour market experience, evaluated at 14 years of experience. According to this estimator, the return to experience decreases from 3.1 percent for the first year to 2.2 percent after 10 years to 1.2 percent after 20 years (see Table II.2). Estimates of Rochina-Barrachina's (1999) most general estimator (the GMM) are slightly lower. They range from 2.2 percent after the first year to 1.4 percent after 10 years to 0.4 percent after 20 years. Simple OLS estimates are intermediate. They range from 3.0 percent after 1 year to 1.9 percent after 10 years to 0.8 percent after 20 years of labour market experience.

³² For Wooldridge's estimator, however, the assumptions under the null hypothesis are stronger than what is

Our results also indicate that estimates of aggregate wage growth are sensitive to the trend in sample selection. If sample selection decreases over time, simple difference estimators lead to downward biased time effects. In our case, wage growth over the 12 years period due to the aggregate time trend is 14 percent for Wooldridge's most general estimator, and 16 percent for Rochina-Barrachina's most general estimator. In contrast, a simple difference estimator assigns only 9 percent of wage growth to aggregate time effects over the 12 years period.

REFERENCES

BLUNDELL, R. and T. MACURDY (1999), "Labour Supply: A Review of Alternative Approaches", in: O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, Vol. III, North-Holland, Amsterdam.

BROWNING, M, DEATON, A., and M. IRISH (1985), "A Profitable Approach to Labour Supply and Commodity Demand over the Life Cycle, *Econometrica*, 53, 503-543.

CARD, D. (1994), "Earnings, schooling and ability revisited", National Bureau of Economic Research, Working Paper 4832.

CHAMBERLAIN, G. (1980), "Analysis of covariance with qualitative data", Review of Economic Studies, 47, 225-238.

CHAMBERLAIN, G. (1984), "Panel data", in Handbook of Econometrics, ed. by Z. Griliches, M.D. Intriligator, Vol. II, 1247-1317, Amsterdam: North-Holland, Ch. 22.

CHARLIER, E., B. MELENBERG AND A. VAN SOEST (1997), "An analysis of housing expenditure using semiparametric models and panel data", CentER Discussion Paper, no. 9714, Tilburg University, The Netherlands.

HECKMAN, J.(1979), "Sample selection bias as a specification error", Econometrica, 47, 153-161.

HOROWITZ, J. (1992), "A smoothed maximum score estimator for the binary response model", Econometrica, 60, 505-531.

HSIAO, C. (1986), Analysis of panel data, Cambridge: Cambridge University Press.

KYRIAZIDOU, E. (1994), "Estimation of a panel data sample selection model", unpublished manuscript, Northwestern University.

KYRIAZIDOU, E. (1997), "Estimation of a panel data sample selection model", Econometrica, 65, 1335-1364.

MaCURDY, T. E. (1981), "An empirical model of labor supply in a life cycle setting", Journal of Political Economy, 89, 1059-1085.

MANSKI, C. (1987), "Semiparametric analysis of random effects linear models from binary panel data", Econometrica, 55, 357-362.

MOFFITT, R. (1984), "Profiles of fertility, labour supply and wages of married women: a complete life-cycle model", Review of Economic Studies, 51, 263-278.

MUNDLACK, Y.(1978), "On the pooling of time series and cross-sectional data", Econometrica, 46, 69-86.

NIJMAN, T. AND M. VERBEEK (1992), "Nonresponse in panel data: the impact on estimates of a life cycle consumption function", Journal of Applied Econometrics, 7, 243-257.

POWELL, J. L. (1994), "Estimation of semiparametric models", Handbook of Econometrics, Vol. 4, 2444-2521.

ROCHINA-BARRACHINA, M. E. (1999), "A new estimator for panel data sample selection models", Annales d'Économie et de Statistique, 55/56, 153-181.

TALLIS, G. M. (1961), "The moment generating function of the truncated multinormal distribution", Journal of the Royal Statistical Society, 23, Series b, 223-229.

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

WAGNER, G., R. BURKHAUSER AND F. BEHRINGER (1993), "The English language public use file of the German Socio-Economic Panel", Journal of Human Resources, 27, 429-433.

ZABEL, J.E.(1992), "Estimating fixed effects and random effects with selectivity", Economics Letters, 40, 269-272.

Appendix I: Econometric Model of Wages

Our econometric model of wages may be motivated as follows. Consider a model where human capital is accumulated in a learning by doing way. The accumulation equation for human capital (measured in monetary units) is then given by:

$$w_{it}^* = w_{it-1}^* + (r_{it-1}d_{it-1})\xi + \left[(r_{it-1}d_{it-1})^2 + \left| \sum_{s=1}^{t-1} r_{is}d_{is}r_{i\tau}d_{i\tau} - \sum_{s=1}^{t-2} r_{is}d_{is}r_{i\tau}d_{i\tau} \right| \right] \zeta. \quad (\text{I.1})$$

Here d_{it} is the participation-status variable and r_{is} is the proportion of time individual i allocates in period s to the labour market. Thus, $(r_{it-1}d_{it-1})$ is equivalent to the increase in human capital in a given period. Human capital depreciates while working, which is reflected by the term in brackets. There is no depreciation in periods out of work. The actual market wage is given by $w_{it} = w_{it}^* + \bar{\alpha}_i + \varepsilon_{it}$, where $\bar{\alpha}_i$ is an individual effect and ε_{it} is some idiosyncratic shock. By recursion, we obtain the following wage equation:

$$w_{it} = w_{i1}^* + \left| \sum_{s=1}^{t-1} r_{is}d_{is} \right| \xi + \left(\sum_{s=1}^{t-1} r_{is}d_{is} \right)^2 \zeta + \bar{\alpha}_i + \varepsilon_{it}, \quad (\text{I.2})$$

where the wage in period t depends on the initial wage, w_{i1}^* , and cumulative work experience and its square.

We assume that the entry wage, w_{i1}^* , is solely determined by the individual's unobserved ability, and the level of schooling:

$$w_{i1}^* = S\beta_S + \alpha_i^*, \quad (\text{I.3})$$

where S is a measure for years of education, and α_i^* is an error term specific to the individual (e.g. "ability"). Combining (I.2) and (I.3) gives:

$$w_{it} = S\beta_S + \sum_{s=1}^{t-1} r_{is} d_{is} \zeta + \left(\sum_{s=1}^{t-1} r_{is} d_{is} \right)^2 \zeta + \alpha_i + \varepsilon_{it}, \quad (\text{I.4})$$

where $\alpha_i \equiv (\bar{\alpha}_i + \alpha_i^*)$. The specification in (5.1) is obtained by using $\sum_{s=1}^{t-1} r_{is} d_{is} = \text{Exp}_{it}$ and by adding to (I.4) time dummies, which reflect aggregate wage growth.

Appendix II: Tables

TABLE II.1: ESTIMATES FOR THE WAGE EQUATION^a

| Variable | (1) OLS | (2) FE | (3) DE (OLS) | (4) DE (IV) | (5) DE (GMM) | (6) ^b W (MD) | (7) ^c W (MD) _(Exp) | (8) ^d K | (9) ^d K (IV) | (10) ^e RB | (11) ^e RB (IV) | (12) ^e RB (GMM) |
|---|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-------------------------------|--|----------------------------|-------------------------------|-----------------------------|---------------------------------|----------------------------------|
| CST | 0.9990* (0.0310) | | | | | 1.1111* (0.0724) | 1.0162* (0.0804) | | | | | |
| D85 | 0.0047 (0.0204) | 0.0056 (0.0139) | 0.0052 (0.0068) | -0.0062 (0.0074) | -0.0041 (0.0029) | 0.0247 (0.0250) | 0.0482* (0.0236) | 0.0052 (0.0068) | 0.0052 (0.0068) | 0.0275 (0.0203) | 0.0249 (0.0200) | -0.0088 (0.0087) |
| D86 | 0.0450* (0.0206) | 0.0308 (0.0163) | 0.0291* (0.0095) | 0.0054 (0.0115) | 0.0168* (0.0038) | 0.0466 (0.0295) | 0.0556* (0.0264) | 0.0291* (0.0095) | 0.0291* (0.0095) | 0.0711* (0.0223) | 0.0536* (0.0226) | 0.0327* (0.0080) |
| D87 | 0.0773* (0.0205) | 0.0588* (0.0199) | 0.0597* (0.0125) | 0.0238 (0.0158) | 0.0332* (0.0046) | 0.0876* (0.0332) | 0.0920* (0.0257) | 0.0597* (0.0125) | 0.0597* (0.0125) | 0.1001* (0.0252) | 0.0960* (0.0267) | 0.0835* (0.0108) |
| D88 | 0.0826* (0.0213) | 0.0492* (0.0240) | 0.0602* (0.0159) | 0.0131 (0.0205) | 0.0317* (0.0056) | 0.1048* (0.0409) | 0.1212* (0.0350) | 0.0602* (0.0159) | 0.0602* (0.0159) | 0.1296* (0.0303) | 0.1253* (0.0326) | 0.0916* (0.0121) |
| D89 | 0.1051* (0.0205) | 0.0614* (0.0284) | 0.0715* (0.0194) | 0.0131 (0.0254) | 0.0341* (0.0066) | 0.1128* (0.0468) | 0.1142* (0.0347) | 0.0715* (0.0194) | 0.0715* (0.0194) | 0.1635* (0.0358) | 0.1437* (0.0398) | 0.1165* (0.0137) |
| D90 | 0.1399* (0.0209) | 0.0941* (0.0330) | 0.1048* (0.0230) | 0.0355 (0.0302) | 0.0568* (0.0077) | 0.1394* (0.0543) | 0.1466* (0.0402) | 0.1048* (0.0230) | 0.1048* (0.0230) | 0.2043* (0.0393) | 0.1863* (0.0447) | 0.1679* (0.0149) |
| D91 | 0.1453* (0.0213) | 0.1142* (0.0378) | 0.1254* (0.0268) | 0.0454 (0.0352) | 0.0622* (0.0089) | 0.1452* (0.0582) | 0.1421* (0.0378) | 0.1254* (0.0268) | 0.1254* (0.0268) | 0.2126* (0.0449) | 0.1872* (0.0505) | 0.1843* (0.0164) |
| D92 | 0.1684* (0.0213) | 0.1274* (0.0426) | 0.1434* (0.0304) | 0.0523 (0.0403) | 0.0766* (0.0104) | 0.1909* (0.0660) | 0.1605* (0.0403) | 0.1434* (0.0304) | 0.1434* (0.0304) | 0.2342* (0.0508) | 0.2227* (0.0568) | 0.1987* (0.0174) |
| D93 | 0.1683* (0.0221) | 0.1258* (0.0475) | 0.1439* (0.0342) | 0.0422 (0.0453) | 0.0706* (0.0109) | 0.1919* (0.0719) | 0.1991* (0.0430) | 0.1439* (0.0342) | 0.1439* (0.0342) | 0.2495* (0.0556) | 0.2308* (0.0626) | 0.2688* (0.0195) |
| D94 | 0.1724* (0.0215) | 0.1276* (0.0525) | 0.1461* (0.0380) | 0.0335 (0.0502) | 0.0628* (0.0112) | 0.2227* (0.0790) | 0.2073* (0.0455) | 0.1461* (0.0380) | 0.1461* (0.0380) | 0.2274* (0.0602) | 0.2361* (0.0670) | 0.2769* (0.0218) |
| D95 | 0.2159* (0.0225) | 0.1398* (0.0572) | 0.1594* (0.0415) | 0.0367 (0.0549) | 0.0668* (0.0131) | 0.2591* (0.0856) | 0.2519* (0.0512) | 0.1594* (0.0415) | 0.1594* (0.0415) | 0.2736* (0.0659) | 0.2795* (0.0748) | 0.3556* (0.0238) |
| ED | 0.1133* (0.0020) | | | | | 0.1065* (0.0042) | 0.1086* (0.0043) | | | | | |
| EXP | 0.0309* (0.0019) | 0.0349* (0.0062) | 0.0324* (0.0042) | 0.0522* (0.0058) | 0.0473* (0.0017) | 0.0230* (0.0090) | 0.0320* (0.0060) | 0.0525* (0.0222) | 0.0157 (0.1935) | 0.0244* (0.0060) | 0.0248* (0.0071) | 0.0229* (0.0021) |
| EXP2 | -0.0058* (0.0005) | -0.0045* (0.0002) | -0.0044* (0.0002) | -0.0065* (0.0003) | -0.0060* (0.0003) | -0.0029* (0.0009) | -0.0049* (0.0012) | -0.0041 (0.0050) | -0.0014 (0.0470) | -0.0041* (0.0005) | -0.0045* (0.0006) | -0.0047* (0.0002) |
| $\partial w/\partial EXP$ (14 years) | 0.0148* (0.0007) | 0.0223* (0.0056) | 0.0200* (0.0039) | 0.0340* (0.0054) | 0.0305* (0.0014) | 0.0148* (0.0077) | 0.0182* (0.0038) | 0.0409* (0.0105) | 0.0116 (0.0637) | 0.0129* (0.0054) | 0.0122* (0.0062) | 0.0097* (0.0017) |
| Av. ret. T. dummies | 0.1204* (0.0150) | 0.0850* (0.0336) | 0.0953* (0.0228) | 0.0268 (0.0301) | 0.0461* (0.0075) | 0.1387* (0.0492) | 0.1399* (0.0306) | 0.0953* (0.0228) | 0.0953* (0.0228) | 0.1739* (0.0380) | 0.1624* (0.0428) | 0.1607* (0.0137) |

^a The numbers in parentheses are standard errors.

^b Standard errors **corrected** for the first stage maximum likelihood probit estimates.

^c Standard errors **corrected** for the first stage maximum likelihood probit estimates and the use of predicted regressors.

^d Standard errors **corrected** for the prior in the time dummies coefficients.

^e Standard errors **corrected** for the first stage maximum likelihood bivariate probit estimates.

*Statistically different from zero at the five-percent significance level.

TABLE II.2: ESTIMATED RATES OF RETURN FOR WORK EXPERIENCE ($\partial w/\partial EXP$)^a

| Years of work experience | (1) OLS | (2) FE | (3) DE (OLS) | (4) DE (IV) | (5) DE (GMM) | (6) ^b W (MD) | (7) ^c W (MD) (Exp) | (8) ^d K | (9) ^d K (IV) | (10) ^e RB | (11) ^e RB (IV) | (12) ^e RB (GMM) |
|--------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-------------------------------|--|----------------------------|-------------------------------|----------------------------|---------------------------------|----------------------------------|
| 1 | 0.0298* (0.0018) | 0.0340* (0.0061) | 0.0315* (0.0041) | 0.0509* (0.0057) | 0.0461* (0.0017) | 0.0224* (0.0089) | 0.0310* (0.0058) | 0.0516* (0.0213) | 0.0154 (0.1842) | 0.0236* (0.0059) | 0.0239* (0.0070) | 0.0220* (0.0021) |
| 5 | 0.0252* (0.0015) | 0.0304* (0.0059) | 0.0280* (0.0041) | 0.0457* (0.0056) | 0.0413* (0.0016) | 0.0201* (0.0085) | 0.0271* (0.0051) | 0.0483* (0.0177) | 0.0143 (0.1468) | 0.0203* (0.0057) | 0.0203* (0.0067) | 0.0182* (0.0019) |
| 10 | 0.0194* (0.0010) | 0.0259* (0.0057) | 0.0236* (0.0040) | 0.0392* (0.0055) | 0.0353* (0.0015) | 0.0172* (0.0080) | 0.0222* (0.0043) | 0.0442* (0.0134) | 0.0128 (0.1003) | 0.0162* (0.0055) | 0.0158* (0.0064) | 0.0135* (0.0018) |
| 15 | 0.0137* (0.0006) | 0.0214* (0.0056) | 0.0192* (0.0039) | 0.0327* (0.0054) | 0.0293* (0.0014) | 0.0143* (0.0076) | 0.0172* (0.0038) | 0.0400* (0.0099) | 0.0113 (0.0547) | 0.0121* (0.0053) | 0.0113 (0.0062) | 0.0088* (0.0017) |
| 20 | 0.0079* (0.0004) | 0.0170* (0.0055) | 0.0148* (0.0039) | 0.0262* (0.0053) | 0.0233* (0.0013) | 0.0114 (0.0074) | 0.0123* (0.0035) | 0.0359* (0.0080) | 0.0099 (0.0184) | 0.0080 (0.0052) | 0.0068 (0.0060) | 0.0041* (0.0016) |

^a The numbers in parentheses are standard errors.

^b Standard errors **corrected** for the first stage maximum likelihood probit estimates.

^c Standard errors **corrected** for the first stage maximum likelihood probit estimates and the use of predicted regressors.

^d Standard errors **corrected** for the prior in the time dummies coefficients.

^e Standard errors **corrected** for the first stage maximum likelihood bivariate probit estimates.

* Statistically different from zero at the five-percent significance level.

Appendix III: The Participation Equation

Results for the participation equation for a selection of estimators are given in Table III.1. The first model is a pooled probit, not taking account of a possible correlation between the explanatory variables and the individual effects. Columns (2) and (3) report results from a specification where individual effects are written as a linear projection on leads and lags of time-varying regressors (see Chamberlain (1984)).³³ The estimation procedure consists of two steps. In the first step cross-equation restrictions are ignored, and the γ_t are estimated by probit for each time period separately. The second step is a minimum distance step. The results in column (2) impose the restriction that $\sigma_t = \sigma$ for $t = 84, \dots, 95$. In column (3), σ_{84} has been normalised to 1, and the remaining variances are estimated.

Finally, in column (4) we present results from a fixed effect logit model, as proposed by Chamberlain (1980). This is the estimator used for the weights in Kyriazidou's (1997) method. Since the scaling is different, only the sign (and the ratios) of the coefficients can be compared with the other 3 models.

The estimates for the time dummies show that female labour force participation increases over the length of the panel. Participation probabilities increase until the age of 30-35 (depending on the specification), and decrease thereafter. An increase in other family income (hhinc) has a negative effect on the participation probability, indicating that leisure is a normal good. The dummy for the husband working has a positive effect on the participation probability, but is insignificant in two out of the four specifications. The number of children in different age groups has a negative effect, where the effect decreases with the age group of the children.

³³The individual effect is written as $\eta_i = z_{i1}\delta_1 + \dots + z_{iT}\delta_T + c_i$, with $c_i \sim N(0, \sigma_c^2)$ and independent of z_i . The $u_i = (u_{i1}, \dots, u_{iT})'$ are assumed to be i.i.d. $N(0, \Sigma)$. Define $\sigma_t = (\tilde{\sigma}_t^2 + \sigma_c^2)^{1/2}$, where $\tilde{\sigma}_t^2$ is the t^{th} diagonal element of Σ . Then

$$P[d_{it} = 1 | z_i] = \Phi \left[\frac{z_{it}\gamma - (z_{i1}\delta_1 + \dots + z_{iT}\delta_T)}{\sigma_t} \right] = \Phi [z_{i1}\gamma_{t1} + \dots + z_{iT}\gamma_{tT}] \quad \text{where}$$

The specification in column 1 does not control for correlated individual specific effects, while specifications in the other columns do. When we compare the first two columns, we observe that the effect of the children variables, and other household income decreases quite substantially, while the effect of education increases. This is consistent with the notion that unobserved ability components which increase the woman's competitiveness in the labour market (and therefore her participation propensity) are negatively correlated with the number of children, but positively correlated with educational achievements. They also seem to be negatively correlated with other household income.

The results in column (3) allow for different variances over time. The coefficient of the constant term is similar in columns (1) and (2) but much smaller (in absolute value) in column (3). To test for the 11 additional restrictions imposed on column (2), relative to column (3), we perform a χ^2 test (see Chamberlain (1984) to test for additional restrictions in minimum distance estimators). The increment in the distance statistic is 146.8201 with a p-value = 0.0002, which clearly leads to rejecting the null hypothesis (the test statistic is χ_{93}^2 distributed)³⁴. We conclude that there are different variances over time for the error term in the selection equation.

$$\gamma_t = \sigma_t^{-1} (\delta'_1, \dots, \delta'_{t-1}, \gamma' - \delta'_t, \delta'_{t+1}, \dots, \delta'_T)'$$

³⁴ The degree of freedom is 104 (the number of parameter estimates in the minimum distance for column (3)) minus 11 (the number of additional restrictions imposed in the minimum distance estimator of column (2)).

TABLE III.1: SOME ESTIMATES FOR THE PARTICIPATION EQUATION^a

| Variables | (1) Pooled probit | (2) Chamberlain(1984) $\sigma = 1$ | (3) Chamberlain (1984, $\sigma_{84} = 1$) | (4) Conditional logit (1980) |
|---------------|----------------------|--|--|------------------------------------|
| CST | -2.1644* (0.2309) | -1.9921* (0.2662) | -1.4615* (0.2686) | |
| D85 | -0.1394* (0.0575) | -0.0570 (0.0599) | -0.1396* (0.0666) | 0.1897 (0.1548) |
| D86 | 0.0633 (0.0579) | 0.0652 (0.0592) | 0.0558 (0.0638) | 1.1056* (0.1722) |
| D87 | 0.0623 (0.0580) | 0.0573 (0.0601) | 0.0001 (0.0560) | 1.5592* (0.1962) |
| D88 | 0.0679 (0.0580) | 0.0770 (0.0603) | -0.0220 (0.0538) | 2.0835* (0.2265) |
| D89 | 0.1329* (0.0581) | 0.1779* (0.0605) | 0.0878 (0.0586) | 2.7388* (0.2614) |
| D90 | 0.2056* (0.0584) | 0.2014* (0.0609) | 0.1073* (0.0595) | 3.5187* (0.3010) |
| D91 | 0.6617* (0.0623) | 0.6045* (0.0644) | 0.9144* (0.1262) | 5.4085* (0.3570) |
| D92 | 0.2786* (0.0589) | 0.2454* (0.0621) | 0.0988* (0.0541) | 4.7497* (0.3867) |
| D93 | 0.3138* (0.0593) | 0.3269* (0.0634) | 0.1008* (0.0537) | 5.3176* (0.4318) |
| D94 | 0.2920* (0.0595) | 0.2786* (0.0644) | 0.1048* (0.0570) | 5.6722* (0.4770) |
| D95 | 0.2649* (0.0599) | 0.2492* (0.0651) | 0.0889 (0.0581) | 6.0183* (0.5228) |
| AGE | 0.1443* (0.0111) | 0.1432* (0.0124) | 0.1097* (0.0146) | |
| AGE2 | -0.0021* (0.0001) | -0.0022* (0.0001) | -0.0017* (0.0002) | -0.0069* (0.0006) |
| ED | 0.0806* (0.0066) | 0.0902* (0.0071) | 0.0878* (0.0090) | |
| CC1 | -0.7635* (0.0368) | -0.5880* (0.0419) | -1.1583* (0.0941) | -1.9587* (0.1079) |
| CC2 | -0.5757* (0.0298) | -0.4361* (0.0369) | -0.5092* (0.0501) | -1.3773* (0.0907) |
| CC3 | -0.2265* (0.0174) | -0.1027* (0.0260) | -0.2053* (0.0302) | -0.3807* (0.0717) |
| HWORK | 0.1032* (0.0372) | 0.0094 (0.0510) | -0.0281 (0.0439) | 0.2923* (0.1355) |
| HHINC | -0.1383* (0.0070) | -0.0430* (0.0085) | -0.0506* (0.0092) | -0.3334* (0.0375) |
| M | -0.3171* (0.0433) | -0.5324* (0.0766) | -0.2983* (0.0680) | -1.5269* (0.1980) |
| σ_{85} | | | 1.1650* (0.1425) | |
| σ_{86} | | | 1.1404* (0.1140) | |
| σ_{87} | | | 0.8926* (0.0799) | |
| σ_{88} | | | 0.7948* (0.0699) | |
| σ_{89} | | | 0.9103* (0.0860) | |
| σ_{90} | | | 0.8675* (0.0871) | |
| σ_{91} | | | 1.9506* (0.2130) | |
| σ_{92} | | | 0.6530* (0.0583) | |
| σ_{93} | | | 0.6313* (0.0551) | |
| σ_{94} | | | 0.7508* (0.0677) | |
| σ_{95} | | | 0.7940* (0.0743) | |

^a The numbers in parentheses are standard errors.

* Statistically different from zero at the five-percent significance level.

IZA Discussion Papers

| No. | Author(s) | Title | Area | Date |
|-----|--|---|------|-------|
| 61 | R. Winkelmann | Immigration: The New Zealand Experience | 1 | 10/99 |
| 62 | A. Thalmaier | Bestimmungsgründe von Fehlzeiten: Welche Rolle spielt die Arbeitslosigkeit? | 3 | 10/99 |
| 63 | M. Ward | Your Everyday, Average Academic | 5 | 10/99 |
| 64 | M. Ward | Salary and the Gender Salary Gap in the Academic Profession | 5 | 10/99 |
| 65 | H. Lehmann J. Wadsworth A. Acquisti | Crime and Punishment: Job Insecurity and Wage Arrears in the Russian Federation | 4 | 10/99 |
| 66 | E. J. Bird H. Kayser J. R. Frick G. G. Wagner | The Immigrant Welfare Effect: Take-Up or Eligibility? | 3 | 10/99 |
| 67 | R. T. Riphahn A. Thalmaier | Behavioral Effects of Probation Periods: An Analysis of Worker Absenteeism | 1/3 | 10/99 |
| 68 | B. Dietz | Ethnic German Immigration from Eastern Europe and the former Soviet Union to Germany: the Effects of Migrant Networks | 1 | 11/99 |
| 69 | M.-S. Yun | Generalized Selection Bias and the Decomposition of Wage Differentials | 6 | 11/99 |
| 70 | I. N. Gang F.L. Rivera-Batiz | Immigrants and Unemployment in the European Community | 1 | 11/99 |
| 71 | L. Goerke | The Wedge | 3 | 11/99 |
| 72 | J. Fersterer R. Winter-Ebmer | Are Austrian Returns to Education Falling Over Time? | 5 | 11/99 |
| 73 | G. S. Epstein S. Nitzan | The Endogenous Determination of Minimum Wage | 3 | 11/99 |
| 74 | M. Kräkel | Strategic Mismatches in Competing Teams | 5 | 12/99 |
| 75 | B. Henry M. Karanassou D. J. Snower | Adjustment Dynamics and the Natural Rate: An Account of UK Unemployment | 1 | 12/99 |
| 76 | G. Brunello M. Giannini | Selective Schools | 5 | 12/99 |
| 77 | C. M. Schmidt | Knowing What Works: The Case for Rigorous Program Evaluation | 6 | 12/99 |

| | | | | |
|----|-------------------------------------|--|-----|-------|
| 78 | J. Hansen R. Wahlberg | Endogenous Schooling and the Distribution of the Gender Wage Gap | 6 | 12/99 |
| 79 | J. S. Earle Z. Sakova | Entrepreneurship from Scratch: Lessons on the Entry Decision into Self-Employment from Transition Economies | 4 | 12/99 |
| 80 | J. C. van Ours J. Veenman | The Netherlands: Old Emigrants – Young Immigrant Country | 1 | 12/99 |
| 81 | T. J. Hatton S. Wheatley Price | Migration, Migrants and Policy in the United Kingdom | 1 | 12/99 |
| 82 | K. A. Konrad | Privacy, time consistent optimal labor income taxation and education policy | 3 | 12/99 |
| 83 | R. Euwals | Female Labour Supply, Flexibility of Working Hours, and Job Mobility in the Netherlands | 1 | 12/99 |
| 84 | C. M. Schmidt | The Heterogeneity and Cyclical Sensitivity of Unemployment: An Exploration of German Labor Market Flows | 1 | 12/99 |
| 85 | S. Pudney M. A. Shields | Gender and Racial Discrimination in Pay and Promotion for NHS Nurses | 5/6 | 12/99 |
| 86 | J.P. Haisken-DeNew C. M. Schmidt | Money for Nothing and Your Chips for Free? The Anatomy of the PC Wage Differential | 5 | 12/99 |
| 87 | T. K. Bauer | Educational Mismatch and Wages in Germany | 1 | 12/99 |
| 88 | O. Bover P. Velilla | Migration in Spain: Historical Background and Current Trends | 1 | 12/99 |
| 89 | S. Neuman | Aliyah to Israel: Immigration under Conditions of Adversity | 1 | 12/99 |
| 90 | H. Lehmann J. Wadsworth | Tenures that Shook the World: Worker Turnover in Russia, Poland and Britain | 4 | 12/99 |
| 91 | M. Lechner | Identification and Estimation of Causal Effects of Multiple Treatments Under the Conditional Independence Assumption | 6 | 12/99 |
| 92 | R. E. Wright | The Rate of Return to Private Schooling | 5 | 12/99 |
| 93 | M. Lechner | An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany | 6 | 12/99 |
| 94 | M. Eichler M. Lechner | An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt | 6 | 12/99 |
| 95 | P. Cahuc A. Zylberberg | Job Protection, Minimum Wage and Unemployment | 3 | 12/99 |
| 96 | P. Cahuc A. Zylberberg | Redundancy Payments, Incomplete Labor Contracts, Unemployment and Welfare | 3 | 12/99 |

| | | | | |
|-----|--|---|-----|-------|
| 97 | A. Barrett | Irish Migration: Characteristics, Causes and Consequences | 1 | 12/99 |
| 98 | J.P. Haisken-DeNew C. M. Schmidt | Industry Wage Differentials Revisited: A Longitudinal Comparison of Germany and USA | 1/5 | 12/99 |
| 99 | R. T. Riphahn | Residential Location and Youth Unemployment: The Economic Geography of School-to-Work-Transitions | 1 | 12/99 |
| 100 | J. Hansen M. Lofstrom | Immigrant Assimilation and Welfare Participation: Do Immigrants Assimilate Into or Out-of Welfare? | 1/3 | 12/99 |
| 101 | L. Husted H. S. Nielsen M. Rosholm N. Smith | Employment and Wage Assimilation of Male First Generation Immigrants in Denmark | 3 | 1/00 |
| 102 | B. van der Klaauw J. C. van Ours | Labor Supply and Matching Rates for Welfare Recipients: An Analysis Using Neighborhood Characteristics | 2/3 | 1/00 |
| 103 | K. Brännäs | Estimation in a Duration Model for Evaluating Educational Programs | 6 | 1/00 |
| 104 | S. Kohns | Different Skill Levels and Firing Costs in a Matching Model with Uncertainty – An Extension of Mortensen and Pissarides (1994) | 1 | 1/00 |
| 105 | G. Brunello C. Graziano B. Parigi | Ownership or Performance: What Determines Board of Directors' Turnover in Italy? | 1 | 1/00 |
| 106 | L. Bellmann S. Bender U. Hornsteiner | Job Tenure of Two Cohorts of Young German Men 1979 - 1990: An analysis of the (West-)German Employment Statistic Register Sample concerning multivariate failure times and unobserved heterogeneity | 1 | 1/00 |
| 107 | J. C. van Ours G. Ridder | Fast Track or Failure: A Study of the Completion Rates of Graduate Students in Economics | 5 | 1/00 |
| 108 | J. Boone J. C. van Ours | Modeling Financial Incentives to Get Unemployed Back to Work | 3/6 | 1/00 |
| 109 | G. J. van den Berg B. van der Klaauw | Combining Micro and Macro Unemployment Duration Data | 3 | 1/00 |
| 110 | D. DeVoretz C. Werner | A Theory of Social Forces and Immigrant Second Language Acquisition | 1 | 2/00 |
| 111 | V. Sorm K. Terrell | Sectoral Restructuring and Labor Mobility: A Comparative Look at the Czech Republic | 1/4 | 2/00 |
| 112 | L. Bellmann T. Schank | Innovations, Wages and Demand for Heterogeneous Labour: New Evidence from a Matched Employer-Employee Data-Set | 5 | 2/00 |
| 113 | R. Euwals | Do Mandatory Pensions Decrease Household Savings? Evidence for the Netherlands | 3 | 2/00 |
| 114 | G. Brunello A. Medio | An Explanation of International Differences in Education and Workplace Training | 2 | 2/00 |

| | | | | |
|-----|--|---|-----|------|
| 115 | A. Cigno F. C. Rosati | Why do Indian Children Work, and is it Bad for Them? | 3 | 2/00 |
| 116 | C. Belzil | Unemployment Insurance and Subsequent Job Duration: Job Matching vs. Unobserved Heterogeneity | 3 | 2/00 |
| 117 | S. Bender A. Haas C. Klose | IAB Employment Subsample 1975-1995. Opportunities for Analysis Provided by the Anonymised Subsample | 7 | 2/00 |
| 118 | M. A. Shields M. E. Ward | Improving Nurse Retention in the British National Health Service: The Impact of Job Satisfaction on Intentions to Quit | 5 | 2/00 |
| 119 | A. Lindbeck D. J. Snower | The Division of Labor and the Market for Organizations | 5 | 2/00 |
| 120 | P. T. Pereira P. S. Martins | Does Education Reduce Wage Inequality? Quantile Regressions Evidence from Fifteen European Countries | 5 | 2/00 |
| 121 | J. C. van Ours | Do Active Labor Market Policies Help Unemployed Workers to Find and Keep Regular Jobs? | 4/6 | 3/00 |
| 122 | D. Munich J. Svejnar K. Terrell | Returns to Human Capital under the Communist Wage Grid and During the Transition to a Market Economy | 4 | 3/00 |
| 123 | J. Hunt | Why Do People Still Live in East Germany? | 1 | 3/00 |
| 124 | R. T. Riphahn | Rational Poverty or Poor Rationality? The Take-up of Social Assistance Benefits | 3 | 3/00 |
| 125 | F. Büchel J. R. Frick | The Income Portfolio of Immigrants in Germany - Effects of Ethnic Origin and Assimilation. Or: Who Gains from Income Re-Distribution? | 1/3 | 3/00 |
| 126 | J. Fersterer R. Winter-Ebmer | Smoking, Discount Rates, and Returns to Education | 5 | 3/00 |
| 127 | M. Karanassou D. J. Snower | Characteristics of Unemployment Dynamics: The Chain Reaction Approach | 3 | 3/00 |
| 128 | O. Ashenfelter D. Ashmore O. Deschênes | Do Unemployment Insurance Recipients Actively Seek Work? Evidence From Randomized Trials in Four U.S. States | 6 | 3/00 |
| 129 | B. R. Chiswick M. E. Hurst | The Employment, Unemployment and Unemployment Compensation Benefits of Immigrants | 1/3 | 3/00 |
| 130 | G. Brunello S. Comi C. Lucifora | The Returns to Education in Italy: A New Look at the Evidence | 5 | 3/00 |
| 131 | B. R. Chiswick | Are Immigrants Favorably Self-Selected? An Economic Analysis | 1 | 3/00 |

| | | | | |
|-----|---|--|-----|------|
| 132 | R. A. Hart | Hours and Wages in the Depression: British Engineering, 1926-1938 | 7 | 3/00 |
| 133 | D. N. F. Bell R. A. Hart O. Hübler W. Schwerdt | Paid and Unpaid Overtime Working in Germany and the UK | 1 | 3/00 |
| 134 | A. D. Kugler G. Saint-Paul | Hiring and Firing Costs, Adverse Selection and Long-term Unemployment | 3 | 3/00 |
| 135 | A. Barrett P. J. O'Connell | Is There a Wage Premium for Returning Irish Migrants? | 1 | 3/00 |
| 136 | M. Bräuningner M. Pannenberg | Unemployment and Productivity Growth: An Empirical Analysis within the Augmented Solow Model | 3 | 3/00 |
| 137 | J.-St. Pischke | Continuous Training in Germany | 5 | 3/00 |
| 138 | J. Zweimüller R. Winter-Ebmer | Firm-specific Training: Consequences for Job Mobility | 1 | 3/00 |
| 139 | R. A. Hart Y. Ma | Wages, Hours and Human Capital over the Life Cycle | 1 | 3/00 |
| 140 | G. Brunello S. Comi | Education and Earnings Growth: Evidence from 11 European Countries | 2/5 | 4/00 |
| 141 | R. Hujer M. Wellner | The Effects of Public Sector Sponsored Training on Individual Employment Performance in East Germany | 6 | 4/00 |
| 142 | J. J. Dolado F. Felgueroso J. F. Jimeno | Explaining Youth Labor Market Problems in Spain: Crowding-Out, Institutions, or Technology Shifts? | 3 | 4/00 |
| 143 | P. J. Luke M. E. Schaffer | Wage Determination in Russia: An Econometric Investigation | 4 | 4/00 |
| 144 | G. Saint-Paul | Flexibility vs. Rigidity: Does Spain have the worst of both Worlds? | 1 | 4/00 |
| 145 | M.-S. Yun | Decomposition Analysis for a Binary Choice Model | 7 | 4/00 |
| 146 | T. K. Bauer J. P. Haisken-DeNew | Employer Learning and the Returns to Schooling | 5 | 4/00 |
| 147 | M. Belot J. C. van Ours | Does the Recent Success of Some OECD Countries in Lowering their Unemployment Rates Lie in the Clever Design of their Labour Market Reforms? | 3 | 4/00 |
| 148 | L. Goerke | Employment Effects of Labour Taxation in an Efficiency Wage Model with Alternative Budget Constraints and Time Horizons | 3 | 5/00 |
| 149 | R. Lalive J. C. van Ours J. Zweimüller | The Impact of Active Labor Market Programs and Benefit Entitlement Rules on the Duration of Unemployment | 3/6 | 5/00 |

| | | | | |
|-----|--|---|-----|------|
| 150 | J. DiNardo K. F. Hallock J.-St. Pischke | Unions and the Labor Market for Managers | 7 | 5/00 |
| 151 | M. Ward | Gender, Salary and Promotion in the Academic Profession | 5 | 5/00 |
| 152 | J. J. Dolado F. Felgueroso J. F. Jimeno | The Role of the Minimum Wage in the Welfare State: An Appraisal | 3 | 5/00 |
| 153 | A. S. Kalwij M. Gregory | Overtime Hours in Great Britain over the Period 1975-1999: A Panel Data Analysis | 3 | 5/00 |
| 154 | Michael Gerfin Michael Lechner | Microeconomic Evaluation of the Active Labour Market Policy in Switzerland | 6 | 5/00 |
| 155 | J. Hansen | The Duration of Immigrants' Unemployment Spells: Evidence from Sweden | 1/3 | 5/00 |
| 156 | C. Dustmann F. Fabbri | Language Proficiency and Labour Market Performance of Immigrants in the UK | 1 | 5/00 |
| 157 | P. Apps R. Rees | Household Production, Full Consumption and the Costs of Children | 7 | 5/00 |
| 158 | A. Björklund T. Eriksson M. Jäntti O. Raalum E. Österbacka | Brother Correlations in Earnings in Denmark, Finland, Norway and Sweden Compared to the United States | 5 | 5/00 |
| 159 | P.- J. Jost M. Kräkel | Preemptive Behavior in Sequential Tournaments | 5 | 5/00 |
| 160 | M. Lofstrom | A Comparison of the Human Capital and Signaling Models: The Case of the Self-Employed and the Increase in the Schooling Premium in the 1980's | 5 | 6/00 |
| 161 | V. Gimpelson D. Treisman G. Monusova | Public Employment and Redistributive Politics: Evidence from Russia's Regions | 4 | 6/00 |
| 162 | C. Dustmann M. E. Rochina-Barrachina | Selection Correction in Panel Data Models: An Application to Labour Supply and Wages | 6 | 6/00 |