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

Christian Dustmann¹ and María Engracia Rochina-Barrachina²

This Version: August 2001

Abstract

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. Not many studies exist that use these methods in practise. We present and compare alternative estimators, and apply them to a typical problem in applied econometrics: the estimation of the wage returns to experience for females. We discuss the assumptions each estimator imposes on the data, and the problems that occur in our applications. This should be particularly useful to practitioners who consider using such estimators in their own application. All 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 this particular measurement error, and non-strict exogeneity can be taken into account within the estimation methods discussed.

¹ We are grateful to Richard Blundell, Arthur van Soest and Frank Windmeijer for useful comments and suggestions. Financial support from the Spanish foundation “Fundación Ramón Areces” to María E. Rochina-Barrachina is gratefully acknowledged. The usual disclaimer applies.

¹ Department of Economics, University College London. Email: c.dustmann@ucl.ac.uk; phone: +44.20.7679.5832.

² Department of Economics, Universidad de Valencia and University College London. Email: erochina@uv.es; phone: +34.96.382.8343.
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 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 that 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.
Wooldridge (1995) has proposed the first estimator we consider. 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 that 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). It also differences out the individual heterogeneity term in the equation of interest, but it imposes distributional assumptions to derive explicitly the selectivity correction term for the difference equation.

In the first part of the paper we describe, in a unified framework, 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. One objective of the paper is to compare alternative estimators, and to show how the methods can be applied in practise. This is done in the second part of the paper, where we apply the three methods to a typical problem in labour economics: to estimate the effect of actual labour market experience on wages of females. Obtaining a consistent estimate of this parameter is a crucial pre-requisite for the analysis of male-female wage (growth) differentials. A large literature is concerned with this problem, but not many papers address in depth the problems arising when estimating this parameter (see England and Farkas 1988 and Polacheck and Kim 1994 for discussions).

The data for our empirical application is drawn from the first 12 waves of the German Socio-Economic Panel (GSOEP). 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.

---

3 Charlier et al. (1997) use a slightly extended version of the estimation technique in Kyriazidou (1997) to deal with estimation of an endogenous switching regression model on housing expenditure with panel data.
We first present results from standard methods, like fixed effects and difference estimators, and the classical Heckman selection estimator. We then apply the before mentioned estimators, and discuss problems which may occur in a typical application as ours.

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 that 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 generally solve the problem. We show how to address this problem within the methods discussed.

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:

\[ w_{it} = x_{it} \beta + \alpha_i + \epsilon_{it}; \quad i = 1, \ldots, N; \quad t = 1, \ldots, T, \]  

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

where \( 1[.] \) is an indicator function, which is equal to one if its argument is true, and zero otherwise. Furthermore, \( \beta \) and \( \gamma \) are unknown parameter vectors, and \( x_{it}, z_{it} \) are vectors of explanatory variables with possibly common elements\(^4\), including both time variant and time invariant variables, and time effects. The \( \alpha_i \) and \( \eta_i \) are unobservable and time invariant individual specific effects\(^5\), which are possibly correlated with \( x_{it} \) and \( z_{it} \). The \( \epsilon_{it} \) 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 \( \beta \).

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 \( \beta \) under the following condition:

---

\(^4\) 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).

\(^5\) We refer to unobserved individual heterogeneity as fixed effects. We consider these parameters as nuisance parameters that may possibly be correlated with the explanatory variables.
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 \( \alpha_i \) is to use some type of difference estimator. Given identification\(^6\), the consistency condition for an estimator using differences across time instead of level equations is given by the following expression:\(^7\)

\[
E(\epsilon_s | x_s, x_t, d_s = d_t = 1) = E(\epsilon_t | x_t, d_t = 1) + E(\epsilon_s | x_t, d_t = 1) = 0, \quad s \neq t, \tag{2.4}
\]

where \( s \) and \( t \) are time periods.

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

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) matches pairs of

\(^6\) For identification we require the matrix \( E[(x_t - x_s) (x_t - x_s) | x_s, x_t, d_t = d_s = 1] \) to be finite and non-singular.

\(^7\) If \( s = t - 1 \), the data is transformed by applying first differencing over time. Other transformations include mean deviation operators.
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 $\alpha_i, \epsilon_{it}$ and the resulting error term in the selection equation (W3 and W4 below).

- **W1:** The regression function of $\eta_i$ 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: $\eta_i = z_i^\prime \delta + c_i$, 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 $(0, \sigma_v^2)$, where $\tilde{z}_i = (x_i, z_i)^\prime$ with $x_i = (x_{it1}, \ldots, x_{itT})$ and $z_i = (z_{i1}, \ldots, z_{iT})$.

- **W3:** The regression function of $\alpha_i$ on $x_i$ and $v_{it}$ is linear. 8

Accordingly, $E(\alpha_i | \tilde{z}_i, v_{it}) = x_i^\prime \psi_1 + \ldots + x_{it}^\prime \psi_T + \phi v_{it}$. 9 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

8 Alternatively, one may assume that $\alpha_i$ depends only on the time average of $x_{it}$ (see Mundlack 1978, Nijman and Veerbeck 1992, and Zabel 1992).
expectation of $\alpha_i$ given $(z_i, d_{it} = 1)$, which is obtained by integrating

$$E(\alpha_i | z_i, v_{it}) = x_i \psi_1 + ... + x_i \psi_T + \phi_y v_{it} \text{ over } v_{it} \leq z_{i1} \gamma_{i1} + ... + z_{it} \gamma_{iT}.$$  

This yields

$$E(\alpha_i | z_i, d_{it} = 1) = x_i \psi_1 + ... + x_i \psi_T + \phi_y E[v_{it} | z_i, d_{it} = 1].$$

• **W4:** $\varepsilon_{it}$ is mean independent of $z_i$ conditional on $v_{it}$ and its conditional mean is linear on $v_{it}$.

Accordingly, $E(\varepsilon_{it}|z_i, v_{it}) = E(\varepsilon_{it}|v_{it}) = \rho_i v_{it}$. Again, as we do not observe $v_{it}$ but the binary selection indicator $d_{it}$, we integrate $E(\varepsilon_{it}|z_i, v_{it}) = \rho_i v_{it}$ over $v_{it} \leq z_{i1} \gamma_{i1} + ... + z_{it} \gamma_{iT}$, resulting in

$$E(\varepsilon_{it}|z_i, d_{it} = 1) = \rho_i E[v_{it} | z_i, d_{it} = 1].$$


$$E(\alpha_i + \varepsilon_{it}|z_i, d_{it} = 1) = E(\alpha_i | z_i, d_{it} = 1) + E(\varepsilon_{it}|z_i, d_{it} = 1) = x_i \psi_1 + ... + x_i \psi_T + \left(\phi_i + \rho_i\right) E[v_{it} | z_i, d_{it} = 1] \quad (2.3')$$

which results in the following model:

$$w_{it} = x_i \psi_1 + ... + x_i \psi_T + x_i \beta + \ell_i H_{it} = z_{it} \gamma_{i1} + ... + z_{it} \gamma_{iT} + \lambda \left( H_{it} / \sigma_i \right) + e_{it}, \quad (2.5)$$

where $\ell_i = \phi_i + \rho_i$, $H_{it} = z_{i1} \gamma_{i1} + ... + z_{iT} \gamma_{iT}$ is the reduced form index in the selection equation for period $t$, and $\lambda \left( H_{it} / \sigma_i \right) = E[v_{it} | z_i, d_{it} = 1]$.

\[9\] The key point for identifying the vector $\beta$ is that, under $v_{it}$ being independent of $z_i$, and the conditional expectation $E(\alpha_i | z_i, v_{it})$ being linear, the coefficients on the $x_{ir}$, $r = 1, ..., T$, are the same regardless of which $V_{it}$ is in the conditioning set.

\[10\] $z_{i1} \gamma_{i1} + ... + z_{iT} \gamma_{iT}$ is the reduced form index for the selection equation in (2.2), once the time-constant unobserved effect $\eta_i$ is specified as in **W1**.
Notice that, since $d_{i\alpha} = 1$ for $r \neq t$ is not included in the conditioning sets of $E(\alpha_i | \bar{z}_i, d_{i\alpha} = 1)$ and $E(e_{i\alpha} | \bar{z}_i, d_{i\alpha} = 1)$, the selection term $E[v_{i\alpha} | \bar{z}_i, d_{i\alpha} = 1]$ is not strictly exogenous in (2.5). The condition, which holds for the new error term in (2.5), is $E(e_{i\alpha} | \bar{z}_i, d_{i\alpha} = 1) = E(e_{i\alpha} | \bar{z}_i, v_{i\alpha} \leq H_{i\alpha}) = 0$. We refer to this as “contemporaneous exogeneity” of the selection term $E[v_{i\alpha} | \bar{z}_i, d_{i\alpha} = 1]$ with respect to $e_{i\alpha}$ in (2.5).

To obtain estimates for $\lambda(\cdot)$, a probit on $H_{i\alpha} = \bar{z}_{i\alpha} \gamma_{1\alpha} + \ldots + \bar{z}_{iT} \gamma_{T\alpha}$ is estimated for each $t$ in the first step. In the second step, Wooldridge (1995) proposes to estimate equation (2.5) either by minimum distance or pooled OLS regression.\(^{11}\) Under the assumptions $W1$-$W4$, the estimator for $\beta$ is consistent. Since dependence between the unobservables in the selection equation, $v_{i\alpha}$, and the unobservables in the main equation, $(e_{i\alpha}, \alpha_i)$, is allowed for, selection may depend not only on the error $e_{i\alpha}$, but also on the unobserved individual effect $\alpha_i$. For time varying variables we can identify $\beta$ under assumption $W3$.

\(^{11}\) In our application we use the more efficient minimum distance estimator. For details about the extension of Wooldridge’s (1995) pooled OLS procedure to minimum distance estimation see Rochina-Barrachina (2000).
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):

\[
E(e_{it} - e_{is} | z_{it}, z_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) = E(e_{it} | z_{it}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) - E(e_{is} | z_{is}, \alpha_i, \eta_i, d_{is} = d_{is} = 1) \equiv \lambda_{its} - \lambda_{ist}
\]

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

\[
\begin{align*}
\lambda_{its} &= \mathbb{E}(e_{it} | z_{it}, \alpha_i, \eta_i, u_{it} \leq z_{it} \gamma - \eta_i; u_{it} \leq z_{it} \gamma - \eta_i) \\
&= \Lambda\left(z_{it} \gamma - \eta_i, z_{it} \gamma - \eta_i; F(e_{it}, u_{it} | z_{it}, \alpha_i, \eta_i)\right)
\end{align*}
\]

\[
\begin{align*}
\lambda_{ist} &= \mathbb{E}(e_{is} | z_{is}, \alpha_i, \eta_i, u_{is} \leq z_{is} \gamma - \eta_i; u_{is} \leq z_{is} \gamma - \eta_i) \\
&= \Lambda\left(z_{is} \gamma - \eta_i, z_{is} \gamma - \eta_i; F(e_{is}, u_{is} | z_{is}, \alpha_i, \eta_i)\right)
\end{align*}
\]

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')), differencing between periods \( s \) and \( t \) will entirely remove the sample selection problem and, at the same time, the time constant individual heterogeneity component.
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:** $(\varepsilon_{it}, \varepsilon_{it}, u_{it}, u_{it})$ and $(\varepsilon_{it}, \varepsilon_{it}, u_{it}, u_{it})$ are identically distributed conditional on $\tilde{z}_{it}, \tilde{z}_{it}, \alpha_i, \eta_i$. That is, $F(\varepsilon_{it}, \varepsilon_{it}, u_{it}, u_{it} | \tilde{z}_{it}, \tilde{z}_{it}, \alpha_i, \eta_i) = F(\varepsilon_{it}, \varepsilon_{it}, u_{it}, u_{it} | \tilde{z}_{it}, \tilde{z}_{it}, \alpha_i, \eta_i)$. 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 condition is imposed on the selection correction function $\Lambda(\cdot)$. This smoothness condition ensures that once K1 holds, $z_{it} \gamma = z_{it} \gamma$ implies $\lambda_{its} = \lambda_{ist}$.

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

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

for individuals satisfying $d_{it} = d_{is} = 1, s \neq t$ and $z_{it} \gamma = z_{it} \gamma$, is consistent. The resulting error $e_{its} = (\varepsilon_{it} - \varepsilon_{is}) - (\lambda_{its} - \lambda_{ist})$ 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 $z_{it} \gamma = z_{it} \gamma$ with probability one, which is not the case if $z_{it}$ contains a continuous variable. To implement the estimator, Kyriazidou

---

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

13 In this model identification of $\beta$ requires $E[(x_{it} - x_{is})^\top (x_{it} - x_{is}) d_t d_s (z_{it} - z_{is}) \gamma] = 0$ to be finite and non-singular. Given that we require support of $(z_{it} - z_{is}) \gamma$ at zero, an exclusion restriction on the set of regressors in $x_{it}$ is required.
(1997) constructs kernel weights, which are a declining function of the distance \(|z_i \gamma - z_s \gamma|\), and estimates pairwise differenced equations by weighted OLS\(^{14}\).

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

### 2.4 Estimation in Differences II: Rochina-Barrachina’s estimator

This estimator is also based on pairwise differencing equation (2.1) for individuals satisfying \(d_s = d_t = 1, s \neq t\). 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 \(\eta_i\) on \(z_i\) is linear\(^{15}\).
- **RB2:** The errors in the selection equation, \(v_i = u_i + c_i\), are normal \(\left(0, \sigma_i^2\right)\) conditional on \(z_i\).
- **RB3:** The errors \([\left[\varepsilon_i - e_i, u_i, v_i\right]]\) are trivariate normally distributed conditional on \(z_i\).

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

\(^{14}\) The estimator is arbitrarily close to root n-consistency depending on the degree of smoothness one is willing to assume for the kernel function.

\(^{15}\) In Rochina-Barrachina (1999), a non-parametric specification of the conditional mean of \(\eta_i\) is also used, where \(E(\eta_i|z_i)\) is left unrestricted.
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 **RB1-RB3**, the resulting estimation equation is given by

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

where $H_{it} = z_{i1} \gamma_{11} + \cdots + z_{i\tau} \gamma_{\tau \tau}$, $\tau = t, s$, are the resulting reduced form indices in the selection equation for periods $t$ and $s$, and $\rho_{ts} = \rho_{(v_i, f_s)(v_i, f_s)}$ is the correlation coefficient between the errors in the selection equation. Furthermore, $\ell_{it}\lambda\left(\frac{H_{it}}{\sigma_t}, \frac{H_{is}}{\sigma_s}, \rho_{ts}\right) + \ell_{is}\lambda\left(\frac{H_{is}}{\sigma_s}, \frac{H_{it}}{\sigma_t}, \rho_{ts}\right)$ is the conditional mean $E(\varepsilon_{it} - \varepsilon_{is}|z_i, d_{it} = d_{is} = 1)$ derived from the three-dimensional normal distribution assumption in **RB3**. The new error term $e_{its} \equiv (\varepsilon_{it} - \varepsilon_{is}) - [\ell_{is}\lambda_{its} + \ell_{st}\lambda_{its}]$ has a conditional expectation $E(e_{its}|z_i, v_{it} \leq H_{it}, v_{is} \leq H_{is}) = 0$. To construct estimates of the $\lambda_{(\cdot)}$ terms the reduced form coefficients $(\gamma_{r}, \gamma_{r})$ will be jointly determined with $\rho_{ts}$, using a bivariate probit for each combination of time periods. The second step is carried out by applying OLS to equation (2.7).

---

16 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 \( \alpha_i \), with respect to the explanatory variables (to allow for individual correlated heterogeneity) and with respect to the random error term \( \nu_{it} \) (to allow for selection that depends on the unobserved effect \( \alpha_i \)). In the other methods, \( \alpha_i \) is differenced out, and selection may therefore depend on \( \alpha_i \) in an arbitrary fashion.

With respect to the assumptions on the functional form of the sample selection effects, Kyriazidou’s (1997) estimator is the most flexible. It 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, normality for the random component of the unobservables in the selection equation. Second, parameterisation of the way \( \eta_i \) depends on the explanatory variables. Third, an assumption about the relationship between the errors in the main equation and the \( \nu_{it} \) in the selection equation. In Wooldridge (1995) joint normality of unobservables in both equations is not needed once a marginal normality assumption for the \( \nu_{it} \) and a linear projection specification for \( \varepsilon_{it} \) on \( \nu_{it} \) are imposed. In Rochina-Barrachina’s (1999) estimator, joint normality is assumed, and linearity between \( \varepsilon_{it} \) and \( \nu_{it} \) results from the joint normality assumption.

Kyriazidou (1997) does not impose any parametric assumption on the distribution of the unobservables in the model, but the conditional exchangeability assumption imposes restrictions on the time series properties of the model, in that it allows for time effects only in the conditional means. 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.

No method imposes explicitly restrictions on the pattern of serial-correlation in the error processes. 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 \((e_{it})\) relates to the time-varying error in the selection equation \((v_{is})\), for \(s \neq t\). Rochina-Barrachina’s (1999) estimator, due to the joint normality assumption \((RB3)\), imposes linearity on the correlation between \(e_{it}\) and \(v_{is}\) for \(s \neq t\), since it includes \(d_{it}, d_{is}\) in the conditioning set.

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.

Kyriazidou’s (1997) estimator is most appealing, as it is the most flexible, imposing least parametric assumptions. However, in particular applications, problems may arise 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 occur in our specific application, as we demonstrate below.
**TABLE 1: COMPARISON OF ESTIMATORS**

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Estimation</th>
<th>Sample selection effects</th>
<th>Distributional assumptions</th>
<th>Specification of conditional means</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wooldridge</strong></td>
<td>Levels</td>
<td>Parameterized</td>
<td>None</td>
<td>Normal $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$, $u_{it}$, $\alpha_i$, $\eta_i$, $\epsilon_{it}$</td>
</tr>
<tr>
<td></td>
<td>Parameterized</td>
<td>None</td>
<td>Normal random component $c_{ij}$</td>
<td>None $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$, $u_{it}$, $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$</td>
</tr>
<tr>
<td></td>
<td>Time diff.</td>
<td>Unspecified</td>
<td>None</td>
<td>None $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$, $u_{it}$, $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$</td>
</tr>
<tr>
<td></td>
<td>Time diff.</td>
<td>Parameterized</td>
<td>None</td>
<td>Normal $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$, $u_{it}$, $\alpha_i$, $\eta_i$, $\epsilon_{it}$, $\nu_{it}$</td>
</tr>
<tr>
<td></td>
<td>Time dummies or time trend</td>
<td>Yes</td>
<td>Yes</td>
<td>Unspecified $d_{it} = 1$</td>
</tr>
<tr>
<td></td>
<td>Time dummies or time trend</td>
<td>Yes</td>
<td>CE$^b$</td>
<td>$d_{it} = d_{it} = 1$, $\tilde{z}<em>{it} \equiv z</em>{it}$, $\gamma$</td>
</tr>
<tr>
<td></td>
<td>Time dummies or time trend</td>
<td>Yes</td>
<td>subject to joint normality</td>
<td>$d_{it} = d_{it} = 1$</td>
</tr>
</tbody>
</table>

*LP denotes the linear projection operator.

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

4.1 Estimation if regressors are non-strictly exogenous

All the estimators above assume strict exogeneity of the regressors. 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, 
\[ E(\epsilon_i | x_i, \alpha_i) = 0. \] 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 \( \alpha_i \). In this case, the set of valid conditioning variables for the linear projection of \( \alpha_i \) on the regressors differs for different time periods – in period \( t \) the conditioning set is the vector \( x'_i \equiv (x_{i1}, \ldots, x_{it}) \). If however the conditioning set changes over time, the coefficients for the leads and lags of the explanatory variables in the linear projection of \( \alpha_i \) will likewise vary over time, thus invalidating \( W3 \). Hence, the condition for \( \beta \) to be separately identified from \( \psi \) (implying that \( \psi_{t1} = \psi_{11}, \ldots, \psi_{tt} = \psi_{Tt}, \ t = 1, \ldots, T \)) does not hold.

One way to deal with this problem 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 \( \beta \), 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 \( (\lambda_{it}) \) 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 procedure\(^{17}\). Let \( z_i \) be the set of instrumental variables. Then the difference \((x_{it} - x_{is})\) fitted by \( z_i \) is \((\hat{x}_{it} - \hat{x}_{is}) = z_i' \left\{ \sum_j z_j z_j' \right\}^{-1} \sum_j z_j (x_{jt} - x_{js})\), and the IV estimator \( b_{IV} \) has the form

\[
  b_{IV} = \left\{ \sum_i (\hat{x}_{it} - \hat{x}_{is})' (x_{it} - x_{is}) d_{it} d_{is} \right\}^{-1} \sum_i (\hat{x}_{it} - \hat{x}_{is})' (w_{it} - w_{is}) d_{it} d_{is} \left[ (z_{it} - z_{is}) \gamma \right] \tag{4.2}
\]

where \( \psi \left[ (z_{it} - z_{is}) \gamma \right] \) 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. This pre-estimation of instruments does not affect the asymptotic distribution of \( b_{IV} \).

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 difficult.

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

\(^{17}\) The IV version of Kyriazidou’s (1997) estimator has been proved to be consistent in Charlier, Melenberg...
\[ b_{\text{GMM}} = \left\{ \sum_i \tilde{x}_{its} \tilde{z}_{its}^\prime \Omega^{-1} \sum_i \tilde{z}_{its} \tilde{x}_{its} \right\}^{-1} \sum_i \tilde{x}_{its} \tilde{z}_{its}^\prime \Omega^{-1} \sum_i \tilde{z}_{its} (w_{it} - w_{it}^\prime), \]  

(4.3)

where \( \tilde{x}_{its} \equiv \left( (x_{it} - x_{is}), \lambda_{its}, \lambda_{ist} \right) \) and \( \tilde{z}_{its} \equiv (z_{it}^\prime, \lambda_{its}, \lambda_{ist}) \). The matrix \( \Omega \) is given by

\[ \Omega = \sum_i \tilde{z}_{its} \tilde{z}_{its}^\prime r_{its}^2, \]

where \( r_{its} = (w_{it} - w_{is}) - (x_{it} - x_{is})b_{IV}^\prime - \left[ \ell_{its}^\prime \lambda_{ists} + \ell_{ist}^\prime \lambda_{ist} \right] \) are the estimated residuals. The \( z_i \) are defined as above, but now the instrument vector for a given pair \((t, s)\), \( \tilde{z}_{its} \), also includes the corresponding sample selection terms \( \lambda_{ists} \) and \( \lambda_{ist} \). By setting \( \Omega = \sum_i \tilde{z}_{its} \tilde{z}_{its}^\prime \) the GMM estimator becomes a simple IV estimator, and estimates can be used as initial estimates for the GMM estimator.

### 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 \( x_{it}^\ast \) be equal to the true variable \( x_{it} \), plus an individual specific error term:

\[ x_{it}^\ast = x_{it} + e_{it}, \]  

(4.4)

and Van Soest (1997).
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:

$$w_{it} = x_{it}^* \Psi_1 + \ldots + x_{it}^* \Psi_T + x_{it}^* \Psi_1 + \ldots + x_{it}^* \Psi_T + x_{it}^* \beta_1 + x_{it}^* \beta_2 + \ell_i \lambda(H_{it} / \sigma_i) + \left[ e_i - (\psi_1 + \ldots + \psi_T + \beta_1)e_i + (\psi_1 + \ldots + \psi_T + \beta_2)e_i^2 - 2(\psi_1 x_{it}^* + \ldots + \psi_T x_{it}^* + \beta_2 x_{it}^*)e_i \right]$$

(4.5)

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 (4.5)). Hence, it is impossible to find instruments which are correlated with the observed regressors, but uncorrelated with the new error term in (4.5).

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 $Z_i$,

$$x_{it} = Z_i \delta_i + s_{it}.$$  \quad (4.6)

Assume that $\delta_i$ is known since it is identified from

$$x_{it}^* = Z_i \delta_i + s_{it} + \epsilon_i.$$  \quad (4.7)

For Wooldridge’s (1995) estimator, substitution of (4.6) into equation (2.5) yields the following expression
where the term in brackets is the new error term, which is a function of the original error term, of linear and quadratic terms in \( s_{it} \), and of cross products \( s_{it}(Z, \delta_t) \). To obtain consistent estimates, we need to assume that \( E(\text{new error term}|Z, \delta_t) \) is a constant that does not vary with \( Z_i \). This holds if the \( Z_i \) are uncorrelated with the original error term in the equation of interest, and if the \( s_{it} \) are independent of \( Z_i \).  

When estimating the model in differences, writing the true regression equation in (2.6) and (2.7) in terms of the observed variables in (4.4) yields:

\[
w_{it} - w_{is} = (x_{it}^* - x_{is}^*)\beta_1 + (x_{it}^{*2} - x_{is}^{*2})\beta_2 + E(e_{it} - e_{is}|i) + \left[ e_{its} - 2\beta_2(x_{it}^* - x_{is}^*)e_i \right]
\]

\[
= (x_{it} - x_{is})\beta_1 + (x_{it}^{*2} - x_{is}^{*2})\beta_2 + E(e_{it} - e_{is}|i) + \left[ e_{its} - 2\beta_2(x_{it} - x_{is})e_i \right]  
\]

(4.8)

where \( E(e_{it} - e_{is}|i) \) is equal to \( E(e_{it} - e_{is}|\bar{Z}_it, Z_{ist}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) \) for Kyriazidou (1997) and to \( E(e_{it} - e_{is}|\bar{Z}_i, d_{it} = d_{is} = 1) \) 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 \( (x_{it} - x_{is}) \). Therefore, differencing eliminates the endogeneity problem due to measurement error, and the IV estimators in section 4.1 can be used to address the problem of non-strict exogenous regressors.

---

18 Independence guaranties not only that the first conditional moment of \( s_{it} \) is equal to zero, but excludes also conditional heteroskedasticity.

19 Since the error term in (4.8) includes \( (x_{it} - x_{is})e_i \), and \( e_i \) is uncorrelated with \( (x_{it} - x_{is}) \),

\[ E(\left( x_{it} - x_{is} \right)e_i | x_{it} - x_{is}) = E(\left( x_{it} - x_{is} \right)e_i | x_{it}^{*2} - x_{is}^{*2}) = 0. \]

20 Notice that this only works for this specific type of measurement error, and if the specification is quadratic in the respective variable. Still, this is likely to cover many applications. As we have pointed out above, the sample design of panel surveys implies that time constant measurement error is frequent. Furthermore, for many applications, quadratic specifications are sufficient.
5. Empirical Model and Data

5.1 Estimation of wage equations for females

There is a large literature that analyses male-female wage differentials (see e.g. Cain, 1986 for a survey, and Blau and Kahn, 1997, for some recent trends). Much of this literature is concerned with establishing the difference in returns to human capital and work history variables between males and females. To obtain an estimate of this parameter requires consistent estimation of the underlying parameters of the wage equation. This is not a trivial task, as selection and individual heterogeneity lead to estimation problems in straightforward regressions. An additional problem arises from the measurement of work experience. Many data sets have no information on actual work experience, and analysts have used potential experience (Age-Education-6) instead. While in some circumstances being an acceptable approximation for males, this measure is likely to overestimate experience for females, thus resulting in underestimated returns.

Some recent studies use data from longitudinal surveys, which provide measures of actual work experience. This variable however adds to the problems in estimating wage equations for females. Work experience is likely to be correlated with unobservables, which determine current wages. Kim and Polacheck (1991), among others, suggest difference estimators to deal with this problem. This leads to consistent estimates only if the selection process is time constant, as we have shown above. Furthermore, since work experience is the accumulation of past participation decisions, it is unlikely to be strictly exogenous in a wage equation – a requirement for consistent estimation of difference estimators.

Our objective in this application is to obtain an estimate of the effect of work experience on wages for females, using the estimators discussed in Section 2. Our empirical analysis is based on data from a twelve-year panel.

We define the wage equation and the participation equation as follows:  

\[ \text{21} \]

---

21 See Appendix I for the derivation of this specification.
where the variable \( d_{it}^* \) is a latent index, measuring the propensity of the individual to participate in the labour market, and \( d_{it} \) is an indicator variable, being equal to one if the individual participates. 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 to education and time dummies, 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 \( d_{it} = 1 \).

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 (see Altonji and Shakotko 1987 and Dustmann and Meghir 2001 for a discussion). Second, sample selection bias. Sample selection occurs if unobservable characteristics affecting the work decision are correlated with the unobservable characteristics affecting the process determining wages. Failure to control for sample selection may lead to incorrect inference regarding the impact of the observables on wages. This problem is particularly severe for females. 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: 

\[
\text{Exp}_{it} = \sum_{s=1}^{t-1} r_{is} d_{is}, \quad \text{where} \quad r_{is} \text{ is the proportion of time individual } i
\]
allocates in period $s$ to the labour market\textsuperscript{22}. 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 the strict exogeneity condition for this variable. Furthermore, given the above formulation, past shocks to wages affect current experience also by altering the weights $r_{it}$. 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 window of the survey should only be marginally affected by miss-measurement, but the pre-sample experience information is likely to suffer considerably from measurement error. As a consequence, the experience variable is measured with error, 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 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 estimation of the wage equations, we use all females that participate in at least two waves.\textsuperscript{23}

\textsuperscript{22} Labour market experience is formed according to $\text{Exp}_{it} = \text{Exp}_{it-1} + r_{it-1}d_{it-1}$, where we obtain by direct substitution $\text{Exp}_{it} = \sum_{s=1}^{t-1} r_{is} d_{is}$.

\textsuperscript{23} To check whether this selection introduces a bias, we compare the means of explanatory variables for the samples excluding, and including females who are only observed once in participation (5861 and 5915 observations respectively). Differences are very small, and never statistically significant.
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. 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, 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 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 0 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. 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.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Total Sample</th>
<th>Work=1 (6802 observations)</th>
<th>Work=1 dropping observations with missing wages (5915)</th>
<th>Work=1 dropping individuals with participation in one year only and observations with missing wages (5861)</th>
<th>Work=0(5834 observations)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work</td>
<td>dummy variable indicating participation of the female (work=1) or no participation (work=0)</td>
<td>0.538</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Lnwage</td>
<td>log gross hourly real wages (1984 West German Marks)</td>
<td>(0.498)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
<td>(0)</td>
</tr>
<tr>
<td>Exp</td>
<td>years-equivalent worked for money after leaving education</td>
<td>14.373</td>
<td>17.661</td>
<td>17.871</td>
<td>17.931</td>
<td>10.541</td>
</tr>
<tr>
<td>Exp2</td>
<td>experience squared and divided by 10</td>
<td>30.231</td>
<td>40.040</td>
<td>40.669</td>
<td>40.861</td>
<td>18.794</td>
</tr>
<tr>
<td>Time</td>
<td>time (year-1900), we also use time dummies for estimation</td>
<td>89.500</td>
<td>89.477</td>
<td>89.460</td>
<td>89.457</td>
<td>89.526</td>
</tr>
<tr>
<td>Age</td>
<td>age of the female in years</td>
<td>42.263</td>
<td>41.259</td>
<td>41.211</td>
<td>41.205</td>
<td>43.434</td>
</tr>
<tr>
<td>Age2</td>
<td>Age of the female squared and divided by 10</td>
<td>188.527</td>
<td>178.988</td>
<td>178.617</td>
<td>178.592</td>
<td>199.650</td>
</tr>
<tr>
<td>Ed</td>
<td>Education of the female measured as years of schooling</td>
<td>10.847</td>
<td>11.057</td>
<td>11.099</td>
<td>11.103</td>
<td>10.602</td>
</tr>
<tr>
<td>Hhinc</td>
<td>Additional real income per month (in thousands)</td>
<td>2.735</td>
<td>2.439</td>
<td>2.398</td>
<td>2.394</td>
<td>3.080</td>
</tr>
<tr>
<td>M</td>
<td>Dummy variable with value 1 if female married and value 0 if not married</td>
<td>0.845</td>
<td>0.793</td>
<td>0.788</td>
<td>0.787</td>
<td>0.905</td>
</tr>
<tr>
<td>hwork(^b)</td>
<td>Dummy variable with value 1 if husband works and value 0 if does not work</td>
<td>0.862</td>
<td>0.877</td>
<td>0.876</td>
<td>0.875</td>
<td>0.846</td>
</tr>
<tr>
<td>cc1</td>
<td>Number of children up to 3 years old in the household</td>
<td>0.117</td>
<td>0.064</td>
<td>0.060</td>
<td>0.059</td>
<td>0.179</td>
</tr>
<tr>
<td>cc2</td>
<td>Number of children between 3 and 6 years old in the household</td>
<td>0.173</td>
<td>0.118</td>
<td>0.110</td>
<td>0.110</td>
<td>0.238</td>
</tr>
<tr>
<td>cc3</td>
<td>Number of children older than 6 years in the household</td>
<td>0.436</td>
<td>0.393</td>
<td>0.370</td>
<td>0.366</td>
<td>0.485</td>
</tr>
</tbody>
</table>

\(^a\)Standard errors in parenthesis.

\(^b\)The reported sample statistics for this variable are conditional on the female being married.
### TABLE 3: STATE FREQUENCIES

<table>
<thead>
<tr>
<th>No. of Years</th>
<th>Participating Individuals</th>
<th>Number of State Changes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
</tr>
<tr>
<td></td>
<td>Changes</td>
<td>Frequency</td>
</tr>
<tr>
<td>0</td>
<td>241</td>
<td>22.89</td>
</tr>
<tr>
<td>1</td>
<td>45</td>
<td>4.27</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>2.75</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>3.80</td>
</tr>
<tr>
<td>4</td>
<td>53</td>
<td>5.03</td>
</tr>
<tr>
<td>5</td>
<td>47</td>
<td>4.46</td>
</tr>
<tr>
<td>6</td>
<td>37</td>
<td>3.51</td>
</tr>
<tr>
<td>7</td>
<td>49</td>
<td>4.65</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>4.65</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
<td>5.60</td>
</tr>
<tr>
<td>10</td>
<td>61</td>
<td>5.79</td>
</tr>
<tr>
<td>11</td>
<td>82</td>
<td>7.79</td>
</tr>
<tr>
<td>12</td>
<td>261</td>
<td>24.79</td>
</tr>
</tbody>
</table>

1053 100 1053 100

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

<table>
<thead>
<tr>
<th>Years</th>
<th>Ratios Work=1/0 in participation sample</th>
<th>number of Work=1 dropping individuals with participation in one year only and observations with missing wages</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>565/468</td>
<td>482</td>
</tr>
<tr>
<td>85</td>
<td>579/474</td>
<td>500</td>
</tr>
<tr>
<td>86</td>
<td>572/481</td>
<td>512</td>
</tr>
<tr>
<td>87</td>
<td>561/492</td>
<td>493</td>
</tr>
<tr>
<td>88</td>
<td>551/502</td>
<td>479</td>
</tr>
<tr>
<td>89</td>
<td>563/490</td>
<td>488</td>
</tr>
<tr>
<td>90</td>
<td>576/477</td>
<td>480</td>
</tr>
<tr>
<td>91</td>
<td>592/461</td>
<td>496</td>
</tr>
<tr>
<td>92</td>
<td>578/475</td>
<td>503</td>
</tr>
<tr>
<td>93</td>
<td>576/477</td>
<td>487</td>
</tr>
<tr>
<td>94</td>
<td>554/499</td>
<td>482</td>
</tr>
<tr>
<td>95</td>
<td>535/518</td>
<td>459</td>
</tr>
<tr>
<td>84-95</td>
<td>6802/5834</td>
<td>5861</td>
</tr>
</tbody>
</table>
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

\[ \frac{\partial w}{\partial EXP} = \xi + 2\zeta EXP, \]  

(6.1)

where we evaluate the expression in (6.1) at 14 years (the sample average). 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.

Columns (1) and (2) present OLS and the standard random effects estimates (RE) respectively, where we allow for time effects, but not for individual heterogeneity that is correlated to the model regressors. The results are very similar and suggest that, evaluated at 14 years of labour market experience, an additional year increases wages by 1.48 and 1.47 percent respectively. If high ability individuals have a stronger labour market attachment than low ability individuals, then these estimates 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 \( \alpha_i \) or the \( \epsilon_{it} \) terms).

In columns (3) and (4), we present estimators that difference out the fixed effects. Column (3) displays standard fixed-effects (within) estimates (FE), and column (4) difference estimates (DE), where all pair differences within time periods per individual are used. Both estimators allow for individual effects correlated with the explanatory variables. A Hausman test of correlation between the regressors and unobserved individual heterogeneity (comparing the RE and FE estimators) leads to rejection of the \( H_0: \beta_{FE} = \beta_{RE} \) (see Table 5). Thus, the upward bias induced by individual fixed effects and any sample selection bias acting through \( \alpha_i \) should be eliminated. Interestingly,

\[ \text{Standard errors of this term are easily derived from the variances and covariances of the parameter estimates for } \xi \text{ and } \zeta. \]

\[ \text{We estimate pooled OLS on 66 pairs corresponding to 25021 observations.} \]
TABLE 5: Marginal Experience Effects, WAGE EQUATION

<table>
<thead>
<tr>
<th></th>
<th>(1) OLS</th>
<th>(2) RE</th>
<th>(3) FE</th>
<th>(4) DE (OLS)</th>
<th>(5) DE (IV)</th>
<th>(6) DE (GMM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \partial \ln \hat{\text{W}}/\partial \text{EXP} ) (14 years)</td>
<td>0.0148* (0.0077)</td>
<td>0.0147* (0.0013)</td>
<td>0.0223* (0.0056)</td>
<td>0.0200* (0.0039)</td>
<td>0.0340* (0.0054)</td>
<td>0.0305* (0.0014)</td>
</tr>
<tr>
<td><strong>Hausman</strong> (Fixed Effects)</td>
<td></td>
<td>\chi^2_{13} = 160.2 (0.000)</td>
<td>\chi^2_{14} = 92.84 (0.000)</td>
<td>\chi^2_{14} = 55.35 (0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Wald Test</strong> (Selection)</td>
<td>\chi^2_{12} = 17.22 (0.1336)</td>
<td>\chi^2_{12} = 17.44 (0.1336)</td>
<td>\chi^2_{12} = 17.44 (0.1336)</td>
<td>\chi^2_{132} = 292.60 (0.000)</td>
<td>\chi^2_{132} = 311.04 (0.000)</td>
<td>\chi^2_{132} = 3859.11 (0.000)</td>
</tr>
<tr>
<td><strong>Hausman</strong> (Exogeneity)</td>
<td>\chi^2_{29} = 6.03 (0.049)</td>
<td>\chi^2_{2} = 5.66 (0.062)</td>
<td>\chi^2_{29} = 46.39 (0.021)</td>
<td>\chi^2_{145} = 433.15 (0.000)</td>
<td>\chi^2_{145} = 1241.19 (0.000)</td>
<td>\chi^2_{145} = 1241.19 (0.000)</td>
</tr>
</tbody>
</table>

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

- Standard errors corrected for the first stage maximum likelihood probit estimates.
- Standard errors corrected for the first stage maximum likelihood probit estimates and the use of predicted regressors.
- Standard errors corrected for the prior in the time dummies coefficients.
- 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 and the standard RE estimations – point estimates for the fixed effect estimator and the difference estimator are 0.022 and 0.020 respectively.

An explanation for this increase 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.

We argued above that experience is not strictly exogenous in the wage level equation if past wage shocks affect current experience levels. In this case, it is endogenous in the difference equation. A common solution to this problem in standard difference estimators is to use instrumental variable techniques. Column (5) and (6) 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. As instruments, we use all leads and lags of the variables in the sample selection equation.

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. A Hausman-type test to compare the IV and GMM estimators with the OLS estimator in differences leads to a rejection of exogeneity of the experience variables in the difference specification.

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 , our instruments are invalid. We now turn to estimation results which take account of a selection
process that operates both through $\varepsilon$ and $\alpha$, and we demonstrate how the problems of measurement error and pre-determinedness can be solved within this framework.\footnote{We have also estimated models using a standard Heckman two steps and full maximum likelihood estimator. Results for the experience effects are nearly identical to the OLS/RE estimators (see Table II in the Appendix). This is not surprising, as these estimators share the same problems than the OLS/RE level estimators.}

6.1 Wooldridge’s estimator

Estimation results for Wooldridge’s (1995) estimator are presented in columns (7) and (8). Following Mundlack (1978) we specify the conditional mean of the individual effects as a linear projection on the within individual means of experience and its square. Results in column (7) are based on the assumption that experience is (strictly) exogenous. Results in column (8) allow for endogeneity by using predictions\footnote{To obtain the predictions for the experience variable (results in column 8), we predict the vector $\left(\text{Exp}_{11}, \ldots, \text{Exp}_{112}, \text{Exp}_{21}^2, \ldots, \text{Exp}_{112}^2\right)$ for all individuals, using leads and lags of the explanatory variables in the selection equation.} for the experience terms. This procedure takes care of both measurement error, and non-strict exogeneity.

The coefficient estimate for Wooldridge’s (1995) estimator is 0.0148 (column 7), which is exactly equal to the OLS result. It is smaller than the fixed effects estimators in columns (3) and (4). To test for sample selection, we have performed a Wald test on the significance of the selection effects, where $H_0: \ell = 0$. 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 $W_3$ is maintained under $H_0$\footnote{See Wooldridge (1995) for details on this point.}. The value for the test statistic is $\chi^2_{12} = 17.22$, 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 $\psi$ coefficients, where $H_0: \psi = 0$. We reject the null hypothesis, which suggests the presence of correlated fixed effects.

In column (8) we use predictions for the experience variables. This leads to a slight increase of the experience coefficient – exactly what we would expect if experience is pre-determined. Hausman-type tests, comparing (7) and (8), reject exogeneity both after
controlling for correlated heterogeneity and sample selection. We perform Wald tests for the estimates in column (8), testing the null hypotheses that $H_0: \ell = 0$ and $H_0: \psi = 0$. Again, we cannot reject the null hypothesis $H_0: \ell = 0$, 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. 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 estimate.$^{30}$

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 $\chi^2$ test leads to rejecting the null hypothesis (with a p-value of 0.0002).

When applying this method to our data, a further problem arises: 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 experience.

$^{30}$ In principle, to estimate the optimal weighting matrix for the minimum distance step requires estimates for the covariance matrix of the estimators for the different pairs of time periods. 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
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. 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. One possible solution is to use information on aggregate wage growth from other sources. To illustrate the estimator, we use here time effects we obtain from the simple difference estimator in column (4).

Estimation results are displayed in columns (9) and (10). Column (9) displays results of simple weighted OLS estimation of equation (2.6). The IV estimates presented in column (10) are obtained by following the procedure described in section 4.1 above.

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 however is affected by current variables (like children(t), hhinc(t), etc.). Current variables should therefore qualify as valid exclusions. We exclude current other household income hhinc(t) from the instrument set for experience.

The estimator in (9) 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, which may be due to the pre-estimated time effects we are using. The estimator in (10) corrects for non-strict exogeneity of the experience variable in the level equation, after accounting for sample selection and individual heterogeneity. Instrumenting reduces the experience effect to 1.2 percent, but the

diagonal matrix.
effect is not statistically significant (which may be due to the smaller effective sample size used for this estimator).

To test for selectivity bias in the simple difference equation, we use a Hausman-type test, comparing the parameter estimates in column (9) with the difference estimator in column (4). The test compares a linear model where selectivity only enters through the fixed effects (column 4), and a model which incorporates more general selectivity effects (Kyriazidou’s estimator in column 9). We then test the assumption of no selectivity bias in the linear panel data model. The test indicates that the null hypothesis of no selectivity bias is rejected.

6.3 Rochina-Barrachina’s estimator

Columns (11)–(13) present estimates, using the method by Rochina-Barrachina (1999). Column (11) displays results of simple OLS estimation of equation (2.7). IV-GMM estimates are presented in columns (12) and (13). 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. 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.

The mean value of the correlation coefficient between the errors in the selection equation in two time periods is 0.7862 (se=0.1299), with a minimum value of 0.4845 and a maximum value of 0.9658. Correlation appears because of the $c_i$ component in the error term and/or because of serially correlated idiosyncratic errors.

To test whether the 66*2 correction terms are jointly significant, we use a Wald test. The values for the test statistics for the estimators in Columns (11) to (13) are clearly larger than the critical values of the $\chi^2_{132}$ at any conventional significance level. Furthermore, Hausman-type tests comparing the IV and the GMM estimators with the OLS estimator in column (11) lead to rejecting exogeneity, after controlling for correlated heterogeneity and sample selection.
The estimated parameters are slightly lower than the OLS/RE 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 has to correct only for the non-strict exogeneity problem. With Wooldridge’s (1995) estimator, the use of predicted regressors corrects also for the measurement error bias, thus leading to larger differences in estimates, compared to the baseline model.

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. Estimates differ according to the methods used. Rochina-Barrachina’s estimator in column (13) assigns about 8 percent more wage growth over the 12 years period to time effects than the simple OLS estimator in column 4. 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 (3) to (6). This leads to downward biased time dummies. With Rochina-Barrachina’s (1999) estimator, the time dummies will presumably pick up the secular productivity growth, since it controls for the decline in sample selection over time.

To investigate whether sample selection does indeed decrease over time, we write the estimated values of the conditional mean \( E(\varepsilon_i - \varepsilon_{is} | \bar{z}_i, d_i = d_{is} = 1), s < t \) 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.\textsuperscript{31}

\textsuperscript{31} This result is in line with the estimates obtained for the participation equation in Appendix III. Here the estimates for the time dummies suggest 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, and the outcome equation contains an unobserved individual specific component that is correlated with the model regressors. In this paper we discuss three estimators that address both problems simultaneously. 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 in the main equation – problems that are likely to occur in many practical applications. We illustrate that the methods of Kyriazidou (1997) and Rochina-Barrachina (1999) can be straightforwardly extended to IV or GMM type estimators. For Wooldridge’s (1995) estimator, we propose to use predicted regressors that are constructed according to the problem at hand.

Not many applications exist for sample selection estimators in panel data models. To understand how the different methods perform in 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 most flexible and attractive estimator is that by Kyriazidou (1997). It avoids specifying the sample selection process, and it requires no parametric assumptions about the unobservables in the model. It does however impose 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, that can only be solved by using additional information. We use pre-estimated time dummies from simple difference estimators. 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 percent. The estimates are clearly
sensitive to the pre-estimated time effects, and it is likely that the simple difference estimator leads to an underestimate of the time effects, as we pointed out in the last section. 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. Using Wooldgridge’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 that 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.

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

---

32 For Wooldridge’s estimator, however, the assumptions under the null hypothesis are stronger than what is required for simple fixed effects.
estimator assigns only 9 percent of wage growth to aggregate time effects over the 12 years period.
REFERENCES


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


Appendix I: The Wage Equation

Assume that human capital accumulation takes the form of learning by doing, and that all human capital is general. Accumulation of human capital (measured in monetary units) is given by:

\[
W_i^* = W_{i-1}^* + \left( r_{i-1}d_{i-1} \right) \xi + \left[ \left( r_{i-1}d_{i-1} \right) \right]^2 + \left( \sum_{x=1}^{t-1} \sum_{s=1}^{x} r_{is}d_{is}r_{is}d_{is} - \sum_{x=1}^{t-2} \sum_{s=1}^{x} r_{is}d_{is}r_{is}d_{is} \right) \zeta - \delta \tag{I.1}
\]

Here \( d_{is} \) is a binary variable indicating labour market participation of individual \( i \) in period \( t \), and \( r_{is} \) is the proportion of time individual \( i \) allocates in period \( s \) to the labour market. Thus, \( \left( r_{i-1}d_{i-1} \right) \) 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. It also depreciates at the rate \( \delta \) during each period of non-work. The actual market wage is given by \( W_{it} = w_i^* + \bar{\alpha}_i + \varepsilon_{it} \), where \( \bar{\alpha}_i \) is an individual specific productivity effect and \( \varepsilon_{it} \) is some idiosyncratic shock. By recursion, we obtain the following wage equation:

\[
W_{it} = W_{i1}^* + \left( \sum_{x=1}^{t-1} r_{ix}d_{ix} \right) \xi + \left( \sum_{x=1}^{t-1} r_{ix}d_{ix} \right)^2 + \left( \sum_{x=1}^{t-2} \sum_{s=1}^{x} r_{is}d_{is}r_{is}d_{is} - \sum_{x=1}^{t-3} \sum_{s=1}^{x} r_{is}d_{is}r_{is}d_{is} \right) \zeta - \delta (t-1) + \bar{\alpha}_i + \varepsilon_{it}, \tag{I.2}
\]

where the wage in period \( t \) depends on the initial wage, \( w_{i1}^* \), cumulative work experience and its square, and the depreciation in periods out of work.

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

\[
w_{i1}^* = S \beta_z + \alpha_z^*, \tag{I.3}
\]

where \( S \) is a measure for years of education, and \( \alpha_z^* \) is an individual specific effect. Combining (I.2) and (I.3) gives:
\[ w_{it} = S\beta_s + \left( \sum_{s=1}^{i-1} r_{is} d_{is} \right) \xi + \left( \sum_{s=1}^{i-1} r_{is} d_{is} \right) \zeta - \delta (t-1) + \alpha_i + \epsilon_{it}, \] (I.4)

where \( \alpha_i \equiv (\alpha_i + \alpha_i^*) \). The specification in (5.1) is obtained by using \( \sum_{s=1}^{i-1} r_{is} d_{is} = \text{Exp}_{it} \) and by adding time dummies, which reflect aggregate wage growth. Notice that the coefficients on the time dummies pick up the aggregate wage growth, net of the average depreciation in skills, due to periods out of work.
### Appendix II: Tables

#### Table II.1: Estimates for the Wage Equation\(^a\)

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1) OLS</th>
<th>(2) RE</th>
<th>(3) FE</th>
<th>(4) DE (OLS)</th>
<th>(5) DE (IV)</th>
<th>(6) DE (GMM)</th>
<th>(7)(^b) W (MD)</th>
<th>(8)(^b) W (MD) (Exp(_2))</th>
<th>(9)(^b) K (MD)</th>
<th>(10)(^b) RB</th>
<th>(11)(^b) RB (IV)</th>
<th>(12)(^b) RB (GMM)</th>
<th>(14)(^b) Heckman (2-steps)</th>
<th>(15)(^b) Heckman (full MLE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CST</td>
<td>0.990*</td>
<td>0.985*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.111*</td>
<td>1.016*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.082*</td>
</tr>
<tr>
<td>D85</td>
<td>0.0047</td>
<td>0.0112</td>
<td>0.0056</td>
<td>0.0052</td>
<td>-0.0062</td>
<td>-0.0041</td>
<td>0.0247</td>
<td>0.0482*</td>
<td>0.0052</td>
<td>0.0275</td>
<td>0.0249</td>
<td>-0.0088</td>
<td>0.0075</td>
<td>0.0078</td>
</tr>
<tr>
<td>D86</td>
<td>0.0450*</td>
<td>0.0433*</td>
<td>0.0308</td>
<td>0.0291*</td>
<td>0.0054</td>
<td>0.0168*</td>
<td>0.0466</td>
<td>0.0556*</td>
<td>0.0291*</td>
<td>0.0711*</td>
<td>0.0536*</td>
<td>0.0327*</td>
<td>0.0436</td>
<td>0.0430*</td>
</tr>
<tr>
<td>D87</td>
<td>0.0773*</td>
<td>0.0774*</td>
<td>0.0588*</td>
<td>0.0597*</td>
<td>0.0238</td>
<td>0.0332*</td>
<td>0.0867*</td>
<td>0.0920*</td>
<td>0.0597*</td>
<td>0.1001*</td>
<td>0.0960*</td>
<td>0.0835*</td>
<td>0.0763*</td>
<td>0.0762*</td>
</tr>
<tr>
<td>D88</td>
<td>0.0826*</td>
<td>0.0749*</td>
<td>0.0492*</td>
<td>0.0602*</td>
<td>0.0131</td>
<td>0.0317*</td>
<td>0.1048*</td>
<td>0.1212*</td>
<td>0.0602*</td>
<td>0.1296*</td>
<td>0.1253*</td>
<td>0.0916*</td>
<td>0.0835*</td>
<td>0.0837*</td>
</tr>
<tr>
<td>D89</td>
<td>0.1051*</td>
<td>0.1009*</td>
<td>0.0614*</td>
<td>0.0715*</td>
<td>0.0131</td>
<td>0.0341*</td>
<td>0.1128*</td>
<td>0.1142*</td>
<td>0.0715*</td>
<td>0.1635*</td>
<td>0.1437*</td>
<td>0.1165*</td>
<td>0.1043*</td>
<td>0.1042*</td>
</tr>
<tr>
<td>D90</td>
<td>0.1399*</td>
<td>0.1337*</td>
<td>0.0941*</td>
<td>0.1048*</td>
<td>0.0355</td>
<td>0.0568*</td>
<td>0.1394*</td>
<td>0.1466*</td>
<td>0.1048*</td>
<td>0.2043*</td>
<td>0.1863*</td>
<td>0.1679*</td>
<td>0.1389*</td>
<td>0.1389*</td>
</tr>
<tr>
<td>D91</td>
<td>0.1453*</td>
<td>0.1592*</td>
<td>0.1142*</td>
<td>0.1254*</td>
<td>0.0454</td>
<td>0.0622*</td>
<td>0.1452*</td>
<td>0.1421*</td>
<td>0.1254*</td>
<td>0.2126*</td>
<td>0.1872*</td>
<td>0.1843*</td>
<td>0.1444*</td>
<td>0.1444*</td>
</tr>
<tr>
<td>D92</td>
<td>0.1684*</td>
<td>0.1794*</td>
<td>0.1274*</td>
<td>0.1434*</td>
<td>0.0523</td>
<td>0.0766*</td>
<td>0.1909*</td>
<td>0.1605*</td>
<td>0.1343*</td>
<td>0.2342*</td>
<td>0.2227*</td>
<td>0.1987*</td>
<td>0.1656*</td>
<td>0.1651*</td>
</tr>
<tr>
<td>D93</td>
<td>0.1683*</td>
<td>0.1839*</td>
<td>0.1258*</td>
<td>0.1439*</td>
<td>0.0422</td>
<td>0.0706*</td>
<td>0.1919*</td>
<td>0.1991*</td>
<td>0.1439*</td>
<td>0.2495*</td>
<td>0.2308*</td>
<td>0.2688*</td>
<td>0.1657*</td>
<td>0.1653*</td>
</tr>
<tr>
<td>D94</td>
<td>0.1724*</td>
<td>0.1917*</td>
<td>0.1276*</td>
<td>0.1461*</td>
<td>0.0335</td>
<td>0.0628*</td>
<td>0.2272*</td>
<td>0.2073*</td>
<td>0.1461*</td>
<td>0.2474*</td>
<td>0.2361*</td>
<td>0.2769*</td>
<td>0.1707*</td>
<td>0.1703*</td>
</tr>
<tr>
<td>D95</td>
<td>0.2159*</td>
<td>0.2119*</td>
<td>0.1398*</td>
<td>0.1594*</td>
<td>0.0367</td>
<td>0.0668*</td>
<td>0.2591*</td>
<td>0.2519*</td>
<td>0.1594*</td>
<td>0.2736*</td>
<td>0.2795*</td>
<td>0.3556*</td>
<td>0.2160*</td>
<td>0.2160*</td>
</tr>
<tr>
<td>ED</td>
<td>0.1133*</td>
<td>0.1125*</td>
<td>0.1065*</td>
<td>0.1086*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP</td>
<td>0.0309*</td>
<td>0.0274*</td>
<td>0.0349*</td>
<td>0.0324*</td>
<td>0.0522*</td>
<td>0.0473*</td>
<td>0.0230*</td>
<td>0.0320*</td>
<td>0.0525*</td>
<td>0.0157</td>
<td>0.0244*</td>
<td>0.0229*</td>
<td>0.0303*</td>
<td>0.0301*</td>
</tr>
<tr>
<td>EXP2</td>
<td>0.0058*</td>
<td>0.0045*</td>
<td>0.0054*</td>
<td>0.0044*</td>
<td>0.0045*</td>
<td>0.0056*</td>
<td>0.0045*</td>
<td>0.0056*</td>
<td>0.0045*</td>
<td>0.0014</td>
<td>0.0045*</td>
<td>0.0045*</td>
<td>0.0005*</td>
<td>0.0005*</td>
</tr>
<tr>
<td>Av. ret. T. &amp; dummies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1204*</td>
<td>0.1243*</td>
<td>0.0850*</td>
<td>0.0953*</td>
<td>0.0268*</td>
<td>0.0461*</td>
<td>0.1387*</td>
<td>0.1399*</td>
<td>0.0953*</td>
<td>0.0953*</td>
<td>0.1739*</td>
<td>0.1624*</td>
<td>0.1607*</td>
<td>0.1197*</td>
<td>0.1195*</td>
</tr>
</tbody>
</table>

\(^a\) 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.

\(^f\) Statistically different from zero at the five-percent significance level.
### TABLE II.2: ESTIMATED RATES OF RETURN FOR WORK EXPERIENCE (\( \frac{\partial w}{\partial \text{EXP}} \))

<table>
<thead>
<tr>
<th>Years of work experience</th>
<th>(1) OLS</th>
<th>(2) RE</th>
<th>(3) FE</th>
<th>(4) DE (OLS)</th>
<th>(5) DE (IV)</th>
<th>(6) DE (GMM)</th>
<th>(7) W (MD)</th>
<th>(8) W (MD) (Exp)</th>
<th>(9) K</th>
<th>(10) K (IV)</th>
<th>(11) RB</th>
<th>(12) RB (IV)</th>
<th>(13) RB (GMM)</th>
<th>(14) Heckman (2-steps)</th>
<th>(15) Heckman (full MLE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0298*</td>
<td>0.0265*</td>
<td>0.0340*</td>
<td>0.0315*</td>
<td>0.0509*</td>
<td>0.0461*</td>
<td>0.0224*</td>
<td>0.0310*</td>
<td>0.0516*</td>
<td>0.0154</td>
<td>0.0236*</td>
<td>0.0239*</td>
<td>0.0220*</td>
<td>0.0292*</td>
<td>0.0290*</td>
</tr>
<tr>
<td></td>
<td>(0.0018)</td>
<td>(0.0022)</td>
<td>(0.0061)</td>
<td>(0.0041)</td>
<td>(0.0057)</td>
<td>(0.0017)</td>
<td>(0.0089)</td>
<td>(0.0058)</td>
<td>(0.0213)</td>
<td>(0.1842)</td>
<td>(0.0059)</td>
<td>(0.0070)</td>
<td>(0.0021)</td>
<td>(0.0032)</td>
<td>(0.0018)</td>
</tr>
<tr>
<td>5</td>
<td>0.0252*</td>
<td>0.0229*</td>
<td>0.0304*</td>
<td>0.0280*</td>
<td>0.0457*</td>
<td>0.0413*</td>
<td>0.0201*</td>
<td>0.0271*</td>
<td>0.0483*</td>
<td>0.0143</td>
<td>0.0203*</td>
<td>0.0203*</td>
<td>0.0182*</td>
<td>0.0247*</td>
<td>0.0246*</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0019)</td>
<td>(0.0059)</td>
<td>(0.0041)</td>
<td>(0.0056)</td>
<td>(0.0016)</td>
<td>(0.0085)</td>
<td>(0.0051)</td>
<td>(0.0177)</td>
<td>(0.1468)</td>
<td>(0.0087)</td>
<td>(0.0067)</td>
<td>(0.0019)</td>
<td>(0.0026)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>10</td>
<td>0.0194*</td>
<td>0.0183*</td>
<td>0.0259*</td>
<td>0.0236*</td>
<td>0.0392*</td>
<td>0.0353*</td>
<td>0.0172*</td>
<td>0.0222*</td>
<td>0.0442*</td>
<td>0.0128</td>
<td>0.0162*</td>
<td>0.0158*</td>
<td>0.0135*</td>
<td>0.0192*</td>
<td>0.0192*</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0016)</td>
<td>(0.0057)</td>
<td>(0.0040)</td>
<td>(0.0055)</td>
<td>(0.0015)</td>
<td>(0.0080)</td>
<td>(0.0043)</td>
<td>(0.0134)</td>
<td>(0.1003)</td>
<td>(0.0055)</td>
<td>(0.0064)</td>
<td>(0.0018)</td>
<td>(0.0019)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>15</td>
<td>0.0137*</td>
<td>0.0138*</td>
<td>0.0214*</td>
<td>0.0192*</td>
<td>0.0327*</td>
<td>0.0293*</td>
<td>0.0143*</td>
<td>0.0172*</td>
<td>0.0400*</td>
<td>0.0113</td>
<td>0.0121*</td>
<td>0.0113</td>
<td>0.0088*</td>
<td>0.0137*</td>
<td>0.0137*</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0013)</td>
<td>(0.0056)</td>
<td>(0.0039)</td>
<td>(0.0054)</td>
<td>(0.0014)</td>
<td>(0.0076)</td>
<td>(0.0038)</td>
<td>(0.0099)</td>
<td>(0.0547)</td>
<td>(0.0053)</td>
<td>(0.0062)</td>
<td>(0.0017)</td>
<td>(0.0013)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>20</td>
<td>0.0079*</td>
<td>0.0093*</td>
<td>0.0170*</td>
<td>0.0148*</td>
<td>0.0262*</td>
<td>0.0233*</td>
<td>0.0114</td>
<td>0.0123*</td>
<td>0.0359*</td>
<td>0.0099</td>
<td>0.0080</td>
<td>0.0068</td>
<td>0.0041*</td>
<td>0.0082*</td>
<td>0.0082*</td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0012)</td>
<td>(0.0055)</td>
<td>(0.0039)</td>
<td>(0.0053)</td>
<td>(0.0013)</td>
<td>(0.0074)</td>
<td>(0.0035)</td>
<td>(0.0080)</td>
<td>(0.0184)</td>
<td>(0.0052)</td>
<td>(0.0060)</td>
<td>(0.0016)</td>
<td>(0.0009)</td>
<td>(0.0006)</td>
</tr>
</tbody>
</table>

* The numbers in parentheses are standard errors.

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

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

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

* 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 model, which does not take 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)). The estimation procedure consists of two steps. In the first step cross-equation restrictions are ignored, and the $\gamma_t$ 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 $t = 84, ..., 95$. In column (3), $\sigma_{84}$ 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 effect of education is positive, indicating that educational achievements increase participation. The number of

---

The individual effect is written as $\eta_i = z_{it}\delta_1 + \ldots + z_{iT}\delta_T + c_i$, with $c_i \sim N(0, \sigma_c^2)$ and independent of $z_i$. The $u_i = (u_{i1}, \ldots, u_{iT})'$ are assumed to be i.i.d. $N(0, \Sigma)$. Define $\sigma_i = (\sigma_i^2 + \sigma_c^2)^{1/2}$, where $\sigma_i^2$ is the $i^{th}$ diagonal element of $\Sigma$. Then

$$P[d_{it} = 1 | z_i] = \Phi\left[\frac{z_{it}'\gamma - (z_{i1}\delta_1 + \ldots + z_{iT}\delta_T)}{\sigma_i}\right] = \Phi\left[z_{i1}'\gamma + \ldots + z_{iT}\gamma'\right]$$

where

$$\gamma_t = \sigma_t^{-1}\left(\delta_1, \ldots, \delta_{t-1}, \gamma' - \delta', \delta_{t+1}, \ldots, \delta_T\right)'$$
children in different age groups has a negative effect, where the effect decreases with the age of the children.

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. This is consistent with the notion that unobserved ability components that increase the woman’s competitiveness in the labour market (and therefore her participation propensity) are negatively correlated with the number of children. 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 $\chi^2$ test. The increment in the distance statistic is 146.82, with a p-value = 0.0002, which clearly leads to rejecting the null hypothesis (the test statistic is $\chi^2_{93}$ distributed). We conclude that there are different variances over time for the error term in the selection equation.
TABLE III.1: THE PARTICIPATION EQUATION

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CST</td>
<td>-2.1644*</td>
<td>-1.9921*</td>
<td>-1.4615*</td>
<td>2.1937*</td>
</tr>
<tr>
<td></td>
<td>(0.2309)</td>
<td>(0.2662)</td>
<td>(0.2686)</td>
<td></td>
</tr>
<tr>
<td>D85</td>
<td>-0.1394*</td>
<td>-0.0570</td>
<td>-0.1396*</td>
<td>0.1897</td>
</tr>
<tr>
<td></td>
<td>(0.0575)</td>
<td>(0.0599)</td>
<td>(0.0666)</td>
<td></td>
</tr>
<tr>
<td>D86</td>
<td>0.0633</td>
<td>0.0652</td>
<td>0.0558</td>
<td>1.056*</td>
</tr>
<tr>
<td></td>
<td>(0.0579)</td>
<td>(0.0592)</td>
<td>(0.0638)</td>
<td></td>
</tr>
<tr>
<td>D87</td>
<td>0.0623</td>
<td>0.0573</td>
<td>0.0001</td>
<td>1.5592*</td>
</tr>
<tr>
<td></td>
<td>(0.0580)</td>
<td>(0.0601)</td>
<td>(0.0560)</td>
<td></td>
</tr>
<tr>
<td>D88</td>
<td>0.0679</td>
<td>0.0770</td>
<td>-0.0220</td>
<td>2.0835*</td>
</tr>
<tr>
<td></td>
<td>(0.0580)</td>
<td>(0.0603)</td>
<td>(0.0538)</td>
<td></td>
</tr>
<tr>
<td>D89</td>
<td>0.1329*</td>
<td>0.1779*</td>
<td>0.0878</td>
<td>2.7388*</td>
</tr>
<tr>
<td></td>
<td>(0.0581)</td>
<td>(0.0605)</td>
<td>(0.0586)</td>
<td></td>
</tr>
<tr>
<td>D90</td>
<td>0.2056*</td>
<td>0.2014*</td>
<td>0.1073*</td>
<td>3.5187*</td>
</tr>
<tr>
<td></td>
<td>(0.0584)</td>
<td>(0.0609)</td>
<td>(0.0595)</td>
<td></td>
</tr>
<tr>
<td>D91</td>
<td>0.6617*</td>
<td>0.6045*</td>
<td>0.9144*</td>
<td>5.4085*</td>
</tr>
<tr>
<td></td>
<td>(0.0623)</td>
<td>(0.0644)</td>
<td>(0.1262)</td>
<td></td>
</tr>
<tr>
<td>D92</td>
<td>0.2786*</td>
<td>0.2454*</td>
<td>0.0988*</td>
<td>4.7497*</td>
</tr>
<tr>
<td></td>
<td>(0.0589)</td>
<td>(0.0621)</td>
<td>(0.0541)</td>
<td></td>
</tr>
<tr>
<td>D93</td>
<td>0.3138*</td>
<td>0.3269*</td>
<td>0.1008*</td>
<td>5.3176*</td>
</tr>
<tr>
<td></td>
<td>(0.0593)</td>
<td>(0.0634)</td>
<td>(0.0537)</td>
<td></td>
</tr>
<tr>
<td>D94</td>
<td>0.2920*</td>
<td>0.2786*</td>
<td>0.1048*</td>
<td>5.6722*</td>
</tr>
<tr>
<td></td>
<td>(0.0595)</td>
<td>(0.0644)</td>
<td>(0.0570)</td>
<td></td>
</tr>
<tr>
<td>D95</td>
<td>0.2649*</td>
<td>0.2492*</td>
<td>0.0889</td>
<td>6.0183*</td>
</tr>
<tr>
<td></td>
<td>(0.0599)</td>
<td>(0.0651)</td>
<td>(0.0581)</td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>0.1443*</td>
<td>0.1432*</td>
<td>0.1097*</td>
<td>0.5228</td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td>(0.0124)</td>
<td>(0.0146)</td>
<td></td>
</tr>
<tr>
<td>AGE2</td>
<td>-0.0021*</td>
<td>-0.0022*</td>
<td>-0.0017*</td>
<td>0.0069*</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0002)</td>
<td></td>
</tr>
<tr>
<td>ED</td>
<td>0.0806*</td>
<td>0.0902*</td>
<td>0.0878*</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td>(0.0066)</td>
<td>(0.0071)</td>
<td>(0.0090)</td>
<td></td>
</tr>
<tr>
<td>CC1</td>
<td>-0.7635*</td>
<td>-0.5880*</td>
<td>-1.1583*</td>
<td>-1.9587*</td>
</tr>
<tr>
<td></td>
<td>(0.0368)</td>
<td>(0.0419)</td>
<td>(0.0941)</td>
<td>(0.1079)</td>
</tr>
<tr>
<td>CC2</td>
<td>-0.5757*</td>
<td>-0.4361*</td>
<td>-0.5902*</td>
<td>-1.3773*</td>
</tr>
<tr>
<td></td>
<td>(0.0298)</td>
<td>(0.0369)</td>
<td>(0.0501)</td>
<td>(0.0907)</td>
</tr>
<tr>
<td>CC3</td>
<td>-0.2265*</td>
<td>-0.1027*</td>
<td>-0.2053*</td>
<td>-0.3807*</td>
</tr>
<tr>
<td></td>
<td>(0.0174)</td>
<td>(0.0260)</td>
<td>(0.0302)</td>
<td>(0.0717)</td>
</tr>
<tr>
<td>HWORK</td>
<td>0.1032*</td>
<td>0.0094</td>
<td>-0.0281</td>
<td>0.2923*</td>
</tr>
<tr>
<td></td>
<td>(0.0372)</td>
<td>(0.0351)</td>
<td>(0.0439)</td>
<td>(0.1355)</td>
</tr>
<tr>
<td>HHINC</td>
<td>-0.1383*</td>
<td>-0.0430*</td>
<td>-0.0506*</td>
<td>-0.3334*</td>
</tr>
<tr>
<td></td>
<td>(0.0070)</td>
<td>(0.0085)</td>
<td>(0.0092)</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>-0.3171*</td>
<td>-0.5324*</td>
<td>-0.2983*</td>
<td>-1.5269*</td>
</tr>
<tr>
<td></td>
<td>(0.0433)</td>
<td>(0.0766)</td>
<td>(0.0680)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\alpha}$ &amp; 1.1650* &amp; (0.1425) &amp; (0.1140)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\beta}$ &amp; 1.1404* &amp; (0.1140) &amp; (0.0679)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\gamma}$ &amp; 0.8926* &amp; (0.0799) &amp; (0.0699)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\delta}$ &amp; 0.9103* &amp; (0.0860) &amp; (0.0871)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\epsilon}$ &amp; 0.8675* &amp; (0.0871) &amp; (0.0871)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\theta}$ &amp; 1.9506* &amp; (0.2130) &amp; (0.0583)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\eta}$ &amp; 0.6530* &amp; (0.0551) &amp; (0.0551)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\zeta}$ &amp; 0.7508* &amp; (0.0677) &amp; (0.0743)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_{\eta}$ &amp; 0.7940* &amp; (0.0743) &amp; (0.0743)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers in parentheses are standard errors.
* Statistically different from zero at the five-percent significance level.