Econometrics Journal (2007), volume **10**, pp. 263–293.
doi: 10.1111/j.1368-423X.2007.00208.x

Selection correction in panel data models: An application to the estimation of females' wage equations

CHRISTIAN DUSTMANN* AND MARÍA ENGRACIA ROCHINA-BARRACHINA†

**Department of Economics, University College London*
E-mail: c.dustmann@ucl.ac.uk

†Department of Economics, Universidad de Valencia
E-mail: erochina@uv.es

First version received: February 2004; final version accepted: December 2006.

Summary 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 not strict exogeneity can be taken into account within the estimation methods discussed.

Key words: sample selection, panel data, wage equations, returns to work experience

1. INTRODUCTION

In many problems of applied econometrics, we only observe data 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 subpopulation is nonrandomly drawn from the overall population, straightforward regression analysis leads to inconsistent estimators. This problem is well known as sample selection bias, and a number of estimators are available which correct for this (see Heckman 1979; 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 the estimation of Frisch demand functions in the consumption and labour supply literature (see, for instance, Browning *et al.* 1985; Blundell and MaCurdy 1999; MaCurdy 1981). Similarly, in many applications, not all variables that ought to be included in the estimation equation are observed. An example is unobserved ability in wage equations (see Card 1994 for details). 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 problem, and a number of straightforward estimators are available (see e.g. Chamberlain 1984; Hsiao 1986; Lee 2002; Wooldridge 2002; Arellano 2003 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. Some estimators have been proposed that deal with both sources of estimation bias. 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 nonrandom selection of the subpopulation over which the equation of interest is defined.

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 individual 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 proposed by Rochina-Barrachina (1999). It also differences out the individual heterogeneity term in the equation of interest, but it imposes distributional assumptions to derive an explicit expression for the selectivity correction term.

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. 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 *et al.* 1988; 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 nonrandomly 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 wage equation.

We first present results from standard methods, like OLS, fixed effects and difference estimators. 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[1] of the explanatory variables. In many typical applications, like the one we use as an illustration, this assumption is likely to

[1] Let the model be $y_t = x_t \beta + u_t, t = 1, \ldots, T$. We define the explanatory variables $\{x_1, \ldots, x_T\}$ as strictly exogenous if $E(u_t | x_1, \ldots, x_T) = 0, t = 1, \ldots, T$ (see also Wooldridge 2002).

be violated. Some recent papers (Honore and Lewbel 2002; Kyriazidou 2001) consider binary and/or selection models with nonexogenous explanatory variables. We suggest a simple extension to the three estimators to relax this assumption in the main equation, maintaining only the strict exogeneity of the regressors in the selection equation.

Another problem that frequently occurs with panel data is measurement error in explanatory variables that are based on biographical information obtained at the start of the panel. This is particularly the case for work history variables. As the error stems mainly from information relating to events before the panel starts, it is constant over the course of the panel. If the affected variables enter the equation of interest in a nonlinear manner, IV estimation does not generally solve the problem. We show how to address this problem for the special case where the variable of interest is included as a second order polynomial (which is likely to cover a range of specifications).

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

2. THE MODEL AND ESTIMATORS

2.1. The model

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

This model can be written as:

$$w_{it} = x_{it}\beta + \alpha_i + \varepsilon_{it};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T, \tag{1}$$

$$d_{it}^* = z_{it}\gamma + \eta_i + u_{it};\ \ d_{it} = 1\left[d_{it}^* > 0\right], \tag{2}$$

where 1[.] is an indicator function, which is equal to one if its argument is true, and zero otherwise. Furthermore, β and γ are unknown parameter (column) vectors, and x_{it}, z_{it} are vectors of explanatory variables with possibly common elements, including both time variant and time invariant variables, and time effects. We assume throughout that there are exclusion restrictions in (1), although this is not required for some of the estimators we will discuss. The α_i and η_i are unobserved and time invariant individual specific effects, which are possibly correlated with x_{it} and z_{it}. The ε_{it} and u_{it} are unobserved disturbances. The variable w_{it} is only observable if $d_{it} =$ 1. The parameter vector we seek to estimate is β.

We assume that panel data is available. Equation (1) could be estimated in levels by pooled ordinary least squares (OLS). A sufficient condition to obtaining consistent estimates of β is:

$$E\left(\alpha_i + \varepsilon_{it}\,|x_{it}, d_{it} = 1\right) = E\left(\alpha_i\,|x_{it}, d_{it} = 1\right) + E\left(\varepsilon_{it}\,|x_{it}, d_{it} = 1\right) = 0,\ \ \forall t. \tag{3}$$

OLS estimates on the selected subsample are inconsistent if selection is nonrandom, and/or if correlated individual heterogeneity is present. In both cases, the conditional expectation in (3) is unequal to zero.

One way to eliminate the individual heterogeneity term α_i is to use some type of difference estimator. Given identification,[2] a sufficient condition for OLS using differences across time to be consistent is:[3]

$$E\,(\varepsilon_{it} - \varepsilon_{is}\,|x_{it}, x_{is}, d_{it} = d_{is} = 1\,) = 0, \qquad s \neq t, \tag{4}$$

where s and t are time periods.

Since condition (4) puts no restrictions on how the selection mechanism or the regressors relate to α_i, differencing equation (1) across time not only eliminates the problem of correlated individual heterogeneity but also any potential selection problem which operates through α_i.

If conditions (3) or (4) are satisfied, the OLS estimator or the difference estimator, respectively, lead to consistent estimates. If conditions (3) and (4) are violated, consistent estimation requires taking account of the selection process. The estimators we describe in the next section take the consistency requirements (3) or (4) as a starting point. The idea of the estimator by Wooldridge (1995) is to derive an expression for the expected value in (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 (4), which is then added as an additional regressor to the differenced equation. The estimator by Kyriazidou (1997) matches pairs of observations for a given individual for whom the conditional expectation in (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 (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 are required. The time dimension allows controlling for individual effects, 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 α_i, ε_{it} and the resulting error term in the selection equation (*W3* and *W4* below).

(1) ***W1:*** *The conditional expectation of η_i given $z_i = (z_{i1}, \ldots, z_{iT})$ is linear.*

Following Chamberlain (1984), Wooldridge (1995) assumes that the conditional expectation of the individual effects in the selection equation is linear in the leads and lags of the observable variables z_{it}: $\eta_i = z_{i1}\delta_1 + \ldots + z_{iT}\delta_T + c_i$, where c_i is a random component independent of everything else.

[2]Sufficient for identification is that the matrix $E[\,(x_t - x_s)'\,(x_t - x_s)d_t d_s]$ is finite and nonsingular for all pairs $\{s,t\}$.

[3]If $s = t - 1$, the data is transformed by applying first differencing over time.

(2) ***W2:*** *The errors in the selection equation, $v_{it} = u_{it} + c_i$, are independent of $\tilde{z}_i$ and normal $(0, \sigma_t^2)$, where $\tilde{z}_i = (x_i, z_i^+)$ with $x_i = (x_{i1}, \ldots, x_{iT})$ and z_i^+ containing the nonoverlapping elements in z_i.*

(3) ***W3:*** *The conditional expectation of α_i given $\tilde{z}_i$ and ν_{it} is linear.*

Accordingly,[4] $E(\alpha_i | \tilde{z}_i, \nu_{it}) = x_{i1}\psi_1 + \ldots + x_{iT}\psi_T + \phi_t \nu_{it}$.[5,6] We do not observe ν_{it}, however, but only the binary selection indicator d_{it}. Therefore, $E(\alpha_i | \tilde{z}_i, \nu_{it})$ has to be replaced by the expectation of α_i given $(\tilde{z}_i, d_{it} = 1)$, which is obtained by integrating $E(\alpha_i | \tilde{z}_i, \nu_{it}) = x_{i1}\psi_1 + \ldots + x_{iT}\psi_T + \phi_t \nu_{it}$ over $\nu_{it} > -H_{it}$, where $H_{it} = z_{i1}\gamma_{t1} + \ldots + z_{iT}\gamma_{tT}$ is the reduced form index for the selection equation in (2), once the time-constant unobserved effect η_i is specified as in *W1* and we have allowed for different σ_t according to *W2*. This yields $E(\alpha_i | \tilde{z}_i, d_{it} = 1) = x_{i1}\psi_1 + \ldots + x_{iT}\psi_T + \phi_t E[\nu_{it} | \tilde{z}_i, d_{it} = 1]$. No other restriction is imposed on the conditional distribution of α_i given $\tilde{z}_i$ and ν_{it}.

(4) ***W4:*** *ε_{it} is mean independent of $\tilde{z}_i$ conditional on ν_{it} and its conditional expectation is linear inν_{it}.*

Accordingly, $E(\varepsilon_{it} | \tilde{z}_i, \nu_{it}) = E(\varepsilon_{it} | \nu_{it}) = \rho_t \nu_{it}$. Again, as we do not observe ν_{it} but the binary selection indicator d_{it}, we integrate $E(\varepsilon_{it} | \tilde{z}_i, \nu_{it}) = \rho_t \nu_{it}$ over $\nu_{it} > -H_{it}$, resulting in $E(\varepsilon_{it} | \tilde{z}_i, d_{it} = 1) = \rho_t E[\nu_{it} | \tilde{z}_i, d_{it} = 1]$.

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

$$E\left(\alpha_i + \varepsilon_{it} \left| \tilde{z}_i, d_{it} = 1\right.\right) = E\left(\alpha_i \left| \tilde{z}_i, d_{it} = 1\right.\right) + E\left(\varepsilon_{it} \left| \tilde{z}_i, d_{it} = 1\right.\right) = x_{i1}\psi_1 + \ldots + x_{iT}\psi_T + (\phi_t + \rho_t) \cdot E\left[\nu_{it} \left| \tilde{z}_i, d_{it} = 1\right.\right] \qquad (3')$$

which results in the following model:

$$w_{it} = x_{i1}\psi_1 + \ldots + x_{iT}\psi_T + x_{it}\beta + \ell_t \lambda(H_{it}) + e_{it}, \qquad (5)$$

where $\ell_t = \phi_t + \rho_t$, and $\lambda(H_{it}) = E[\nu_{it} | \tilde{z}_i, d_{it} = 1]$.[7] The error term $e_{it} \equiv (\alpha_i + \varepsilon_{it}) - E(\alpha_i + \varepsilon_{it} | \tilde{z}_i, d_{it} = 1)$ has the conditional expectation $E(e_{it} | \tilde{z}_i, d_{it} = 1) = 0$.

Notice that, under assumption *W2* and since ν_{ir} for $r \neq t$ is not included in the conditioning sets of assumptions *W3* and *W4*, the selection term $E[\nu_{it} | \tilde{z}_i, d_{it} = 1]$ is not strictly exogenous in (5).

To obtain estimates for $\lambda(\cdot)$, a probit on H_{it} is estimated for each t in the first step. In the second step, Wooldridge (1995) proposes to consistently estimate equation (5) either by minimum distance or pooled OLS regression.[8] Under the assumptions *W1–W4*, the estimator for β is consistent. Since dependence between the unobservables in the selection equation,ν_{it}, and

[4] Recall that we assume throughout exclusion restrictions on (1). For this reason, even if we condition on $\tilde{z}_i$ the conditional expectation depends only on x_i. This accounts for z_i^+ being independent of α_i and ε_{it}.

[5] Alternatively, one may assume that α_i depends only on the time average of x_{it} (see Mundlak 1978; Nijman and Verbeek 1992; Zabel 1992).

[6] The key point for identifying the vector β is that, under ν_{it} being independent of $\tilde{z}_i$, and the conditional expectation $E(\alpha_i | \tilde{z}_i, \nu_{it})$ being linear, the coefficients on the $x_{ir}, r = 1, \ldots, T$, are the same regardless of which ν_{it} is in the conditioning set (see Wooldridge 1995).

[7] $\lambda(H_{it}) = \phi(H_{it})/\Phi(H_{it})$, where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

[8] In our application we use the minimum distance approach. It is computationally easier to estimate each wave separately by OLS and to impose cross-equation restrictions by minimum distance. For details on the asymptotic distribution and variance for the minimum distance estimator of Wooldridge's (1995) panel data sample selection model see weblink in acknowledgements.

the unobservables in the main equation, $(\varepsilon_{it}, \alpha_i)$, is allowed for, selection may depend not only on the error ε_{it}, but also on the unobserved individual effect α_i. For time varying variables we can identify β under assumption *W3*.

2.3. Estimation in differences

2.3.1. Kyriazidou's estimator. The estimator developed by Kyriazidou (1997) relies on pairwise differences over time applied to model (1) for individuals satisfying $d_{it} = d_{is} = 1$, $s \neq t$. The idea of the estimator is as follows. Reconsider first the expression in (4):

$$\begin{aligned} &\mathrm{E}(\varepsilon_{it} - \varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) \\ &\quad = \mathrm{E}(\varepsilon_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) - \mathrm{E}(\varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, d_{it} = d_{is} = 1) \\ &\quad \equiv \lambda_{its} - \lambda_{ist}, \end{aligned} \tag{4'}$$

where $\tilde{z}_{it} = (x_{it}, z_{it}^{+})$, $\tilde{z}_{is} = (x_{is}, z_{is}^{+})$, with z_{it}^{+}, z_{is}^{+} containing the nonoverlapping elements in z_{it} and z_{is}, respectively, and for each time period the selection terms are

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

$$\begin{aligned} \lambda_{ist} &= \mathrm{E}\left(\varepsilon_{is} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i, u_{is} > -z_{is}\gamma - \eta_i, u_{it} > -z_{it}\gamma - \eta_i\right) \\ &= \Lambda[-z_{is}\gamma - \eta_i, -z_{it}\gamma - \eta_i; F(\varepsilon_{is}, u_{is}, u_{it} | \tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i)] \end{aligned}$$

where $\Lambda(\cdot)$ is an unknown function and $F(\cdot)$ is an unknown joint conditional distribution function of the errors. The 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 [i.e. $\lambda_{its} = \lambda_{ist}$ in equation (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 individual effects.

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

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

This 'conditional exchangeability' assumption may be rather restrictive in practical applications. Necessary conditions are that the marginal distributions of $(\varepsilon_{it}, \varepsilon_{is})$ and (u_{it}, u_{is}) are identical, respectively. As Lee (2002) points out, in view of the necessary conditions, exchangeability seems plausible in stationary environments with weak (the error is generated by a process whose mean and variance are not changing over time) as well as strong (the distribution is not changing over time) stationarity. Such stationary environments are unlikely to hold in many applications, as we demonstrate below. Furthermore, joint conditional stationarity is weaker than joint conditional exchangeability (see Kyriazidou 1997).

For individuals satisfying $d_{it} = d_{is} = 1$ $(s \neq t)$ and $z_{it}\gamma = z_{is}\gamma$, under assumption *K1* and provided identification is met,[9] the OLS estimator applied to

$$w_{it} - w_{is} = (x_{it} - x_{is})\,\beta + (\varepsilon_{it} - \varepsilon_{is})\,, \tag{6}$$

is consistent.[10] In practice, it is difficult to find individuals with $z_{it}\gamma = z_{is}\gamma$ and thus, more generally, (6) becomes $w_{it} - w_{is} = (x_{it} - x_{is})\beta + (\lambda_{its} - \lambda_{ist}) + \vartheta_{its}$, where the error term $\vartheta_{its} \equiv (\varepsilon_{it} - \varepsilon_{is}) - (\lambda_{its} - \lambda_{ist})$ has a conditional expectation that satisfies $E\,(\vartheta_{its}\,|\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i\;, d_{it} = d_{is} = 1) = 0$.

To implement the estimator, Kyriazidou (1997) constructs kernel weights, which are a declining function of the distance $|z_{it}\gamma - z_{is}\gamma|$, and estimates pairwise differenced equations by weighted OLS.[11]

The procedure requires estimates of γ, which can be obtained either by smoothed conditional maximum score estimation (see, for instance, Charlier *et al.* 1997; Kyriazidou 1997) or conditional logit estimation (see Chamberlain 1980).[12]

2.3.2. Rochina-Barrachina's estimator. This estimator is also based on pairwise differencing equation (1) for individuals satisfying $d_{it} = d_{is} = 1$, $s \neq t$. Different from Kyriazidou's (1997) estimator, Rochina-Barrachina's (1999) estimator relies on a parameterisation of the conditional expectation in (4). On the other hand, it does not impose the 'conditional exchangeability' assumption.

To implement the estimator, the following assumptions are made:

(1) **RB1:** *The conditional expectation of η_i given z_i is linear.*[13]
(2) **RB2:** *The errors in the selection equation, $\nu_{it} = u_{it} + c_i$, are independent of $\tilde{z}_i$ and normal* $(0, \sigma_t^2)$.
(3) ***RB3:*** *The errors $[(\varepsilon_{it} - \varepsilon_{is}), \nu_{it}, \nu_{is}]$ are trivariate normally distributed and independent of $\tilde{z}_i$.*

The first two assumptions refer to the selection equation and are similar 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 nonparametric with respect to the individual effects in the main equation and allows, under its semiparametric version, for a nonparametric 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

$$w_{it} - w_{is} = (x_{it} - x_{is})\,\beta + \ell_{ts}\lambda\,(H_{it}, H_{is}, \rho_{ts}) + \ell_{st}\lambda\,(H_{is}, H_{it}, \rho_{ts}) + \xi_{its}, \tag{7}$$

[9] In this model identification of β requires $E[\,(x_t - x_s)'\,(x_t - x_s)d_t d_s \mid (z_t - z_s)\gamma = 0]$ to be finite and non-singular. This shows that an exclusion restriction on the set of regressors in x_{it} is necessary.

[10] The asymptotic distribution and variance of the minimum distance estimator for the Kyriazidou's (1997) panel data model with more than two time periods is derived by Charlier *et al.* (1997). For more details see the website in acknowledgements.

[11] Due to the nonparametric matching involved by this approach convergence of this estimator will be slower than the usual root *n* rate.

[12] In cases with 'small' fixed effects, an alternative class of estimators is suggested by Laisney and Lechner (2003).

[13] We use here the more parametric version of the estimator proposed in Rochina-Barrachina (1999), where the conditional expectation of η_i given z_i is parameterised. Alternatively, Rochina-Barrachina (1999) proposes an estimator where the conditional expectation is left unrestricted.

where $H_{i\tau} = z_{i1}\gamma_{\tau 1} + \ldots + z_{iT}\gamma_{\tau T}$, $\tau = t, s$, are the resulting reduced form indices in the selection equation for periods t and s, and $\rho_{ts} = \rho_{(\nu_t/\sigma_t)(\nu_s/\sigma_s)}$ is the correlation coefficient between the errors in the selection equation. Furthermore, $\ell_{ts}\lambda\ (H_{it}, H_{is}, \rho_{ts}) + \ell_{st}\lambda\ (H_{is}, H_{it}, \rho_{ts})$ is the conditional mean $E(\varepsilon_{it} - \varepsilon_{is}\,|\tilde{z}_i, d_{it} = d_{is} = 1)$ derived from the three-dimensional normal distribution assumption in *RB3*.[14] The new error term $\xi_{its} \equiv (\varepsilon_{it} - \varepsilon_{is}) - [\ \ell_{ts}\lambda_{its} + \ell_{st}\lambda_{ist}]$ has a conditional expectation $E(\xi_{its}\,|\tilde{z}_i, \nu_{it} > -H_{it}, \nu_{is} > -H_{is}) = 0$. To construct estimates of the $\lambda\ (\cdot)$ terms the reduced form coefficients (γ_t, γ_s) will be jointly determined with ρ_{ts}, using a bivariate probit for each combination of time periods. The second step is carried out by applying OLS to equation (7).[15]

3. COMPARISON OF ESTIMATORS

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

With respect to the assumptions on the functional form of the sample selection effects, Kyriazidou'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. Secondly, parameterisation of the way η_i depends on the explanatory variables. Thirdly, an assumption about the relationship between the errors in the main equation and the ν_{it} in the selection equation. In Wooldridge (1995) joint normality of unobservables in both equations is not needed once a conditional mean independence assumption (*W4*), a linear projection specification for ε_{it} on ν_{it} and a marginal normality assumption for the ν_{it} are imposed. In Rochina-Barrachina's (1999) estimator, joint normality is assumed, and linearity between ε_{it} and ν_{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 mean of the regression equation. In Wooldridge (1995) and Rochina-Barrachina (1999) not only the conditional mean of the regression equation, but also the conditional means of the selection equation and 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

[14] In particular, $\lambda\,(H_{it}, H_{is}, \rho_{ts}) = \phi\,(H_{it})\,\Phi\left(M^*_{its}\right)/\Phi_2\,(H_{it}, H_{is}, \rho_{ts})$ and $\lambda\,(H_{is}, H_{it}, \rho_{ts}) = \phi\,(H_{is})\,\Phi\left(M^*_{ist}\right)/\Phi_2\,(H_{it}, H_{is}, \rho_{ts})$, where $M^*_{its} = (H_{is} - \rho_{ts}H_{it})/\left(1 - \rho^2_{ts}\right)^{1/2}$, $M^*_{ist} = (H_{it} - \rho_{ts}H_{is})/\left(1 - \rho^2_{ts}\right)^{1/2}$, $\phi(\cdot)$ is the standard normal density function, and $\Phi(\cdot)$, $\Phi_2(\cdot)$ are the standard univariate and bivariate normal cumulative distribution functions, respectively. See Rochina-Barrachina (1999) for more details.

[15] For the asymptotic distribution and variance of the minimum distance estimator for the Rochina-Barrachina's (1999) panel data model with more than two time periods see the website under acknowledgements.

Table 1. Comparation of estimators

Estimators	Estimation	Sample selection effects	Distributional assumptions				Specification of conditional means		
			α_i	η_i	ε_{it}	u_{it}	α_i	η_i	ε_{it}
Wooldridge	Levels	Parameterised	None	Normal random component c_i	None	Normal	LP[(a)] on x_i & $\nu_{it} = c_i + u_{it}$	LP[(a)] on z_i	CMI[(c)] and LP[(a)] on ν_{it}
Kyriazidou	Time diff.	Unspecified	None	None	None but CE[(b)]	None but CE[(b)]	None	None	None
Rochina-Barrachina	Time diff.	Parameterised	None	Normal random component c_i	Normal	Normal	None	LP[(a)] on z_i/non-parametric	Linearity from joint normality

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

Notes: [(a)]LP denotes the linear projection operator. [(b)]Subject to the 'conditional exchangeability' (CE) assumption according to which the vectors of errors $(\varepsilon_{it}, \varepsilon_{is}, u_{it}, u_{is})$ and $(\varepsilon_{is}, \varepsilon_{it}, u_{is}, u_{it})$ are identically distributed conditional on $\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i$. [(c)]'Conditional mean independence' (CMI) assumption according to which $E(\varepsilon_{it} | \tilde{z}_i, \nu_{it}) = E(\varepsilon_{it} | \nu_{it})$.

on the way the time-varying error in the main equation (ε_{it}) relates to the time-varying error in the selection equation (ν_{is}), for $s \neq t$. Rochina-Barrachina's (1999) estimator, due to the joint normality assumption (*RB3*), imposes linearity on the correlation between ε_{it} and ν_{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'.

The three methods assume i.i.d. cross-section observations. Although Kyriazidou's method does not require in principle an i.i.d. assumption across individuals, but identical distribution over time for a given individual, the asymptotic properties of her estimator are developed under the i.i.d. assumption.

Kyriazidou's (1997) estimator imposes the fewest 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 the indices $z_{it}\gamma$ 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. Similarly, if high matching weights are assigned to observations whose x variables change in a systematic manner, and low matching weights to observations where x changes nonsystematically (for example changes in the number of children is likely to lead to nonsystematic changes in actual experience, thus leading to indices that are not close), it may not be 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. Finally, Kyriazidou's estimator is computationally rather demanding (it is bandwidth dependent and it requires smoothing) and slower than $\sqrt{n}$-consistent.

4. EXTENSIONS

4.1. Estimation if regressors in the main equation are not 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. The above three estimators can be extended to take account of not strict exogeneity in the main equation. We maintain the strict exogeneity assumption of regressors in the selection equation.[16]

As pointed out by Wooldridge (1995), a more complex case arises when the variables in the equation of interest are predetermined, and possibly correlated with the individual effects α_i. In this case, the set of valid conditioning variables for the conditional expectation of α_i given the regressors differs for different time periods, in period t the conditioning set is the vector $x_i^t \equiv (x_{i1}, \ldots, x_{it})$. If however the conditioning set changes over time, the coefficients for the leads and lags of the explanatory variables in the linear projection of α_i will likewise vary over time, thus invalidating *W3*. Hence, the condition for β to be separately identified from ψ (implying that $\psi_{t1} = \psi_1, \ldots, \psi_{tT} = \psi_T, t = 1, \ldots, T$) does not hold.

[16] See Lewbel (2003), Honore and Lewbel (2002) and Kyriazidou (2001) for more general discussions of nonexogenous explanatory variables in binary/selection models.

One way to deal with this problem is to substitute the not strictly exogenous (predetermined) time-varying correlated regressors by their predictions, and to apply Wooldridge´s (1995) estimator. Predictions for this method should be done in a particular way, given that the equations for all time periods require T unique predictions for each not strictly exogenous variable. This accounts for the coefficients of all the leads and lags of the linear projection of α_i on $\hat{x}_i$, where $\hat{x}_i$ is the vector of predicted regressors for the original x_i, to be constant over time. To identify β, assumption *W3* must hold and this constraints the way of obtaining predictions. One way to obtain valid and 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 may 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 $(\hat{x}_{it} - \hat{x}_{is}) = z_i\{\sum_j z_j' z_j\}^{-1} \sum_j z_j'(x_{jt} - x_{js})$, and the IV estimator b_{IV} has the form

$$b_{IV} = \left\{ \sum_i (\hat{x}_{it} - \hat{x}_{is})' (x_{it} - x_{is}) d_{it} d_{is} \varpi_{ts} [(z_{it} - z_{is}) \hat{\gamma}] \right\}^{-1} \times \sum_i (\hat{x}_{it} - \hat{x}_{is})' (w_{it} - w_{is}) d_{it} d_{is} \varpi_{ts} [(z_{it} - z_{is}) \hat{\gamma}], \tag{8}$$

where $\varpi_{ts} [(z_{it} - z_{is}) \hat{\gamma}]$ is the kernel weight for individual i in pair (t, s). This approach allows maintaining 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 nonparametric 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.

A way to allow for endogenous regressors in the main equation for Rochina-Barrachina's (1999) estimator is to use a generalised method of moments estimator of the form

$$b_{GMM} = \left\{ \sum_i \ddot{x}_{its}' \ddot{z}_{its} \Omega^{-1} \sum_i \ddot{z}_{its}' \ddot{x}_{its} \right\}^{-1} \sum_i \ddot{x}_{its}' \ddot{z}_{its} \Omega^{-1} \sum_i \ddot{z}_{its}' (w_{it} - w_{is}), \tag{9}$$

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

[17]The IV version of Kyriazidou's (1997) estimator has been proved to be consistent in Charlier *et al.* (1997).

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 nonlinear way, differencing over time does not eliminate the measurement error. In what follows, we show how to address this problem for the special case where the variable of interest is included as a second order polynomial (which is likely to cover a range of specifications).

For illustration, suppose that the scalar variable x_{it} is measured with error, and we include its level and its square among the regressors in equation (1). Let the measured variable x_{it}^* be equal to the true variable x_{it}, plus an individual specific error term:

$$x_{it}^* = x_{it} + \mu_i, \tag{10}$$

where μ_i is assumed to be uncorrelated with x_{it}, and independent of the errors in the selection equation (the v_{it} in Wooldridge and Rochina-Barrachina's estimators or the u_{it} in Kyriazidou's estimator). Assume that the variable x_{it} enters the equation of interest as a second order polynomial. For Wooldridge's (1995) estimator, writing the true regression equation in (5) in terms of the observed variables leads to the following expression:

$$\begin{aligned} w_{it} = {} & x_{i1}^*\psi_1 + \ldots + x_{iT}^*\psi_T + x_{i1}^{*2}\theta_1 + \ldots + x_{iT}^{*2}\theta_T + x_{it}^*\beta_1 + x_{it}^{*2}\beta_2 + \ell_t\lambda\left(H_{it}\right) \\ & + \left[e_{it} - (\psi_1 + \ldots + \psi_T + \beta_1)\,\mu_i + (\theta_1 + \ldots + \theta_T + \beta_2)\,\mu_i^2 \right. \\ & \left. - 2\left(\theta_1 x_{i1}^* + \ldots + \theta_T x_{iT}^* + \beta_2 x_{it}^*\right)\mu_i\right] \end{aligned} \tag{11}$$

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 no longer lead to consistent estimates in a nonlinear error in variables problem, because the error of measurement is no longer additively separable from the regressors (see expression (11)). Hence, it is impossible to find instruments which are correlated with the observed regressors, but uncorrelated with the new error term in (11).

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

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

$$x_{it} = z_i\delta_t + s_{it}. \tag{12}$$

Assume that δ_t is known since it is identified from

$$x_{it}^* = z_i\delta_t + s_{it} + \mu_i. \tag{13}$$

For Wooldridge's (1995) estimator, substitution of (12) into equation (5) yields the following expression

$$\begin{aligned} w_{it} = &(z_i\delta_1)\psi_1 + \ldots + (z_i\delta_T)\psi_T + (z_i\delta_1)^2\theta_1 + \ldots + (z_i\delta_T)^2\theta_T \\ &+(z_i\delta_t)\beta_1 + (z_i\delta_t)^2\beta_2 + \ell_t\lambda(H_{it}) \\ &+\big[e_{it} + (s_{i1}\psi_1 + \ldots + s_{iT}\psi_T + s_{it}\beta_1) + \big(s_{i1}^2\theta_1 + \ldots + s_{iT}^2\theta_T + s_{it}^2\beta_2\big) \\ &+2((z_i\delta_1)s_{i1}\theta_1 + \ldots + (z_i\delta_T)s_{iT}\theta_T + (z_i\delta_t)s_{it}\beta_2)\big], \end{aligned}$$

where the term in brackets is the new error term, which is a function of the error term (e_{it}) in (5), of linear and quadratic terms in s_{it}, and of cross products $s_{it}(z_i\ \delta_t)$. To obtain consistent estimates of slope parameters, we need to assume that $E(\text{new error term}|z_i\ \delta_t)$ is a constant that does not vary with z_i. This holds if the z_i are uncorrelated with the error term (e_{it}) in (5), and if the s_{it} are independent of z_i. Independence guaranties not only that the first conditional moment of s_{it} is equal to zero, but also excludes conditional heteroskedasticity of s_{it}.

When estimating the model in differences, writing the true regression equation in (6) and (7) in terms of the observed variables in (10) yields:

$$\begin{aligned} w_{it} - w_{is} &= \left(x_{it}^* - x_{is}^*\right)\beta_1 + \left(x_{it}^{*2} - x_{is}^{*2}\right)\beta_2 + E\left(\varepsilon_{it} - \varepsilon_{is}\,|\cdot\right) + \left[\pi_{its} - 2\beta_2\left(x_{it}^* - x_{is}^*\right)\mu_i\right] \\ &= \left(x_{it} - x_{is}\right)\beta_1 + \left(x_{it}^{*2} - x_{is}^{*2}\right)\beta_2 + E\left(\varepsilon_{it} - \varepsilon_{is}\,|\cdot\right) + \left[\pi_{its} - 2\beta_2\left(x_{it} - x_{is}\right)\mu_i\right] \end{aligned} \tag{14}$$

where $E(\varepsilon_{it} - \varepsilon_{is}|\cdot)$ and π_{its} are equal to $E\left(\varepsilon_{it} - \varepsilon_{is}\,|\tilde{z}_{it}, \tilde{z}_{is}, \alpha_i, \eta_i\ , d_{it} = d_{is} = 1\right)$ and ϑ_{its}, respectively, for Kyriazidou (1997) and to $E\left(\varepsilon_{it} - \varepsilon_{is}\,|\tilde{z}_i, d_{it} = d_{is} = 1\right)$ and ξ_{its} for Rochina-Barrachina (1999). The new error is given by the term in brackets. Therefore, differencing does not take care of the measurement error, which will lead to biased and inconsistent OLS estimates. However, the IV estimators in Section 4.1 can be used to address not only the problem of not strictly exogenous regressors but also the measurement error problem.[18]

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, Blau and Kahn 1997 for some recent trends, and Fitzenberger and Wunderlich 2002 for a detailed descriptive analysis for Germany). 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 an easy 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

[18] Notice that this holds only 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.

overestimate experience for females, thus resulting in underestimated returns to labour market experience.

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.

Our objective in this application is to obtain an estimate of the effect of work experience on log wages for females, using the estimators discussed in Section 2. Our empirical analysis is based on data from a 12-year panel.

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

$$w_{it} = x_{it}\beta + Exp_{it}\varphi + Exp_{it}^2\zeta + \alpha_i + \varepsilon_{it};\ \ i = 1, \ldots, N;\ \ t = 1, \ldots, T, \tag{15}$$

$$d_{it}^* = z_{it}\gamma + \eta_i + u_{it};\ \ d_{it} = 1\left[d_{it}^* > 0\right], \tag{16}$$

where w_{it} are log real wages. 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 (15) 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 (15) results in upward biased coefficients (see Altonji and Shakotko 1987; Dustmann and Meghir 2005 for a discussion). Second, sample selection bias through unobservable characteristics affecting the work decision being correlated with unobservable characteristics affecting the process determining wages. This problem is particularly severe for females. Third, experience is likely to be not strictly exogenous, even after controlling for heterogeneity and sample selection. Labour market experience in any period t is an accumulation of weighted past participation decisions: $Exp_{it} = \sum_{s=1}^{t-1} r_{is}d_{is}$, where r_{is} is the proportion of time individual i allocates in period s to the labour market.[19] 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_{is}. A final problem is measurement

[19] Labour market experience is formed according to $Exp_{it} = Exp_{it-1} + r_{it-1}\, d_{it-1}$, where we obtain by direct substitution $Exp_{it} = \sum_{s=1}^{t-1} r_{is}d_{is}$.

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 missmeasurement, 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 and 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 nonparticipants for the estimation of the selection equation. For estimation of the wage equations, we use all females that participate in at least two waves.[20]

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- 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- and full-time work.[21] 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 renegotiated 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- or full-time employment at the interview time. Nonparticipation is defined as being in the state of nonemployment or unemployment. On average, 54% 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 nonworking sample.

We do not restrict our sample to married females. From the 12636 observations, 10680 (84.52%) are married, of whom 51% participate in the labour market. We observe a higher percentage of labour market participants (72%) among the nonmarried. 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.

[20]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.

[21]This assumes that part- and full-time experience adds to human capital in the same way. An alternative would be to give a lower weight to part-time experience. The choice of the weight is problematic, however.

Table 2. Description of variables and sample statistics (12 636 observations)[(a)]

Variable	Description	Total Sample	Work = 1 (6802 observations)	Work = 1 dropping observations with missing wages (5915)	Work = 1 dropping individuals with participation in one year only and observations with missing wages (5861)	Work = 0(5834 observations)
Work	Dummy variable indicating participation of the female (work = 1) or no participation (work = 0)	0.538 (0.498)	1 (0)	1 (0)	1 (0)	0 (0)
Lnwage	Log gross hourly real wages (1984 West German Marks)	2.681 (0.435)	2.681 (0.435)	2.682 (0.431)	2.685 (0.432)	
Exp	Years-equivalent worked for money after leaving education	14.373 (9.782)	17.661 (9.407)	17.871 (9.345)	17.931 (9.331)	10.541 (8.765)
Exp2	Experience squared and divided by 10	30.231 (36.606)	40.040 (38.264)	40.669 (38.083)	40.861 (38.122)	18.794 (30.862)
Age	Age of the female in years	42.263 (9.953)	41.259 (9.356)	41.211 (9.373)	41.205 (9.381)	43.434 (10.487)
Age2	Age of the female squared and divided by 10	188.527 (84.624)	178.988 (76.917)	178.617 (76.890)	178.592 (76.952)	199.650 (91.567)
Ed	Education measured as years of schooling	10.847 (1.958)	11.057 (2.129)	11.099 (2.129)	11.103 (2.128)	10.602 (1.705)

Table 2. (Continued).

Variable	Description	Total Sample	Work = 1 (6802 observations)	Work = 1 dropping observations with missing wages (5915)	Work = 1 dropping individuals with participation in one year only and observations with missing wages (5861)	Work = 0(5834 observations)
Hhinc	Additional real income per month	2.735	2.439	2.398	2.394	3.080
	(in thousands)	(1.778)	(1.855)	(1.895)	(1.897)	(1.617)
M	Dummy variable; 1 if married	0.845	0.793	0.788	0.787	0.905
		(0.361)	(0.404)	(0.409)	(0.409)	(0.293)
hwork[b]	Dummy variable; 1 if husband	0.862	0.877	0.876	0.875	0.846
	works	(0.345)	(0.328)	(0.330)	(0.331)	(0.361)
cc1	Number of children up to 3	0.117	0.064	0.060	0.059	0.179
	years old in the household	(0.399)	(0.301)	(0.288)	(0.287)	(0.481)
cc2	Number of children between 3	0.173	0.118	0.110	0.110	0.238
	and 6 years old in the household	(0.442)	(0.364)	(0.350)	(0.351)	(0.511)
cc3	Number of children older than 6	0.436	0.393	0.370	0.366	0.485
	years in the household	(0.739)	(0.696)	(0.680)	(0.675)	(0.784)

Notes: [a]Standard deviations in parenthesis. [b]The reported sample statistics for this variable are conditional on the female being married.

Table 3. State frequencies

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

Our children variables distinguish between the number of children aged between 0 and 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 nonparticipants.

To estimate our wage equation conditional on individual 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 nonparticipation. Twenty-three precent of our sample individuals participate in none of the 12 years, and about 25% 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 seven times over the 12-years period. In the longitudinal dimension, 767 women (corresponding to 6757 observations) worked for a wage at least in 2 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 4).[22]

6. ESTIMATION RESULTS

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

$$\partial w \big/ \partial Exp = \varphi + 2\zeta Exp, \tag{17}$$

[22] To check whether this selection introduces a bias, we compare the means of explanatory variables for the two samples (6757 and 5861 observations, respectively). Differences are very small, and never statistically significant.

Table 4. Number of observations work = 1 versus work = 0

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

where we evaluate the expression in (17) at 14 years (the sample average).[23] We report estimates of φ and ζ as well as the total effect in (17) in Table 5. The full set of results is given in Table A.1 in Appendix A. Rates of return implied by the different methods and for increasing levels of work experience are presented in Table A.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%, respectively. If high ability individuals have a stronger labour market attachment than low ability individuals, then these estimates should be upward biased. Sample selection, on the other side, may lead to a downward bias if selection is positive, and if participation is positively related to past employment. Likewise, measurement error in the experience variable leads to downward biased estimates.

In columns (3) and (4), we present estimators that difference out the individual 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.[24] A Hausman test of correlation between the regressors and unobserved individual heterogeneity (comparing the RE and FE estimators) leads to rejection of the $Ho : \beta_{FE} = \beta_{RE}$ (see Table 5).

The FE-DE 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,

[23]Standard errors of this term are easily derived from the variances and covariances of the parameter estimates for φ and ζ.

[24]We estimate pooled OLS on 66 pairs corresponding to 25 021 observations.

Table 5. Marginal experience effects, wage equation[(a)]

	(1) OLS	(2) RE	(3) FE	(4) DE (OLS)	(5) DE (IV)	(6) DE (GMM)
Exp	**0.0309***	**0.0274***	**0.0349***	**0.0324***	**0.0522***	**0.0473***
	(0.0019)	(0.0023)	(0.0062)	(0.0042)	(0.0058)	(0.0017)
Exp2	**−0.0058***	**−0.0045***	**−0.0045***	**−0.0044***	**−0.0065***	**−0.0060***
	(0.0005)	(0.0005)	(0.0005)	(0.0002)	(0.0003)	(0.0001)
$\partial w / \partial Exp$	**0.0148***	**0.0147***	**0.0223***	**0.0200***	**0.0340***	**0.0305***
(14 years)	(0.0007)	(0.0013)	(0.0056)	(0.0039)	(0.0054)	(0.0014)
Hausman[1]			$\chi^2_{13} = 160.2$			
(Fixed Effects)			(0.000)			
Hausman[2]					$\chi^2_{14} = 92.84$	$\chi^2_{14} = 55.35$
(Exogeneity and absence of measurement error –ME–)					(0.000)	(0.000)

respectively. One explanation is that the partial elimination in the FE-DE estimates of the downward bias by sample selection working through the fixed effect overcompensates the upward bias in the OLS-RE estimates via the correlated individual effects.

The FE-DE estimates can still have a problem of measurement error. Further, 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 solution to these problems 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.[25] 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 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 the existence of measurement error and/or 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/Wu type test[26] 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 and/or to a rejection of not existence of measurement error.

One concern may be the validity of our instruments as, in particular the children variables. We have re-estimated the IV equations excluding the children variables from the instrument set

[25]The IV estimates are used as the first step estimates to obtain the GMM estimates.

[26]In contrast to the standard form of the Hausman test, the test here does not assume efficiency of one estimator under the null hypothesis. This requires estimation of the covariance matrix between estimators (Lee 1996).

Table 5. (Continued).

	$(7)^{(b)}$ W (MD)	$(8)^{(c)}$ W (MD) $(Ex\hat{p})$	$(9)^{(d)}$ K	$(10)^{(d)}$ K (IV)	$(11)^{(e)}$ RB	$(12)^{(e)}$ RB (GMM)
Exp	**0.0230***	**0.0320***	**0.0525***	**0.0157**	**0.0244***	**0.0229***
	(0.0090)	(0.0060)	(0.0222)	(0.1935)	(0.0060)	(0.0021)
Exp2	**−0.0029***	**−0.0049***	**−0.0041**	**−0.0014**	**−0.0041***	**−0.0047***
	(0.0009)	(0.0012)	(0.0050)	(0.0470)	(0.0005)	(0.0002)
$\partial w / \partial Exp$	**0.0148***	**0.0182***	**0.0409***	**0.0116**	**0.0129***	**0.0097***
(14 years)	(0.0077)	(0.0038)	(0.0105)	(0.0637)	(0.0054)	(0.0017)
Wald Test[3]	$\chi^2_{12} = 17.22$	$\chi^2_{12} = 17.44$			$\chi^2_{132} = 292.60$	$\chi^2_{132} = 3859.11$
(Selection)	(0.1412)	(0.1336)			*(0.000)*	*(0.000)*
Wald Test[4]	$\chi^2_{2} = 6.03$	$\chi^2_{2} = 5.66$				
(Fixed Effects)	(0.049)	(0.062)				
Hausman[5]			$\chi^2_{2} = 6.6332$			
(Selection)			(0.036)			
Hausman[6]		$\chi^2_{29} = 46.39$				$\chi^2_{2} = 3.27$
(Exogeneity and absence of ME)		(0.021)				(0.195)

Notes: [(a)]The numbers in parentheses below the coefficient estimates are standard errors. The numbers in parentheses below the test statistics are p-values. [(b)]Standard errors corrected for the first stage maximum likelihood probit estimates. [(c)]Standard errors corrected for the first stage maximum likelihood probit estimates and the use of predicted regressors. [(d)]Standard errors corrected for the use of pre-estimated time dummies coefficients obtained from the simple difference estimator in (4). [(e)]Standard errors corrected for the first stage maximum likelihood bivariate probit estimates. *Statistically different from zero at the five-percent significance level. [1–6]Please see Appendix B for notes 1–6.

for experience, and including higher polynomials of age and education as well as cross terms of these two variables. The estimates of the rate of return to experience, evaluated at 14 years, are 0.028 and 0.031 for the IV and GMM estimators in columns (5) and (6), respectively. These are close to those we report in the table, and suggest that the differences between these estimators and those obtained in columns (3) and (4) are not due to using children as instruments.

The IV difference estimates are consistent under the assumption that selection only works through the individual effects. We now turn to estimation results which take account of a selection process that operates both through ε and α.[27]

[27]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 Tables A.1 and A.2 in Appendix A). This is not surprising, as these estimators share the same problems as 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 Mundlak (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[28] for the experience terms. This procedure takes care of both measurement error, and not strict exogeneity.

The coefficient estimate for Wooldridge's (1995) estimator is 0.0148 (column 7), which is nearly identical to the OLS/RE results. It is smaller than the standard fixed effects estimators in columns (3) and (4) and the IV difference estimators in columns (5) and (6). To test for sample selection, we have performed a Wald test on the joint significance of the twelve selection effects involved, where *Ho*: $\ell_{84} = 0, \ell_{85} = 0, \ldots, \ell_{95} = 0$. 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, as *W3* is maintained under *Ho*.[29] The value for the test statistic is $\chi^2_{12} = 17.22$, with a *p*-value of 0.1412. Thus, the null hypothesis cannot be rejected. We also performed a Wald test for the joint significance of the ψ coefficients, where *Ho*: $\psi = 0$. We reject the null hypothesis, which suggests the presence of correlated individual effects.

In column (8) we use predictions for the experience variables. This leads to an increase of the experience coefficient—what we would expect if experience is pre-determined and/or there exists measurement error. 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 *Ho*: $\ell = 0$ and *Ho*: $\psi = 0$. Again, we cannot reject the null hypothesis *Ho*: $\ell = 0$, but we reject the null hypothesis *Ho*: $\psi = 0$ at a 6.21% 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). 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. For bandwidth selection in the kernel weights we follow the procedure in Kyriazidou (1997). 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 seems plausible in stationary environments. The assumption that the error terms in the selection equation are stationary over time is testable. We have estimated the selection equation under the assumption of equal variances over time, and allowing for variances to differ across time periods,

[28] To obtain the predictions for the experience variables, we predict the vector $(Exp_{i1}, \ldots, Exp_{i12}, Exp^2_{i1}, \ldots, Exp^2_{i12})$ using the entire sample of individuals in the participation equation, and all leads and lags of the explanatory variables in that equation as instruments.

[29] See Wooldridge (1995) for details on this point.

[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 *et al.* (1997) proof that these covariances converge to zero due to the fact that the bandwidth tends to zero as the sample size increases. As a consequence, the optimal weighting matrix simplifies to a block diagonal matrix.

using Chamberlain´s (1984) estimator (results not reported). A χ^2 test leads to rejecting the null hypothesis of equal variances (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 identifying instruments, like, for instance, the number of children. Any systematic increase in experience between two periods cannot be distinguished from the time trend; any nonsystematic 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 cannot 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 (6). The IV estimates presented in column (10) are obtained by following the procedure described in Section 4.1 above.

Given the nonparametric 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 other household income. Participation in the current period however is affected by current variables (like current other household income). Current variables should therefore qualify as valid exclusions. We exclude current other household income from the instrument set for experience.

The estimator in (9) does not correct for possible endogeneity of the experience variable and/or for measurement error. The coefficient for the experience effect indicates that a year of labour market experience increases wages by 4.1%. This estimate is very large, which may be due to inaccuracies in the pre-estimated time effects we are using. The estimator in (10) corrects both for not strict exogeneity of the experience variable and for measurement error in the main equation, after accounting for sample selection and individual heterogeneity. Instrumenting reduces the experience effect to 1.2%, but the effect is not statistically significant.

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 individual 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) and (12) present estimates, using the method by Rochina-Barrachina (1999). Column (11) displays results of simple OLS estimation of equation (7). GMM estimates are presented in column (12). For estimation, we use each combination of panel waves (t,s), resulting in a total of 66 pairs. To combine these estimates, we use minimum distance. 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 (12) are clearly larger than the critical values of the χ^2_{132} at any conventional significance level.[31] Furthermore, Hausman-type tests comparing the GMM parameter estimates for experience with the OLS estimates in column (11) do not lead to jointly rejecting exogeneity and the absence of measurement error, after controlling for correlated heterogeneity and sample selection. Consequently, RB estimates 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%.

The estimated parameters are lower than the simple difference estimators. Compared to Wooldridge's (1995) estimator, estimates are also smaller, which may be due to different parametric assumptions imposed by the two estimators.

7. CONCLUSIONS

In this paper we discuss three estimators that address the problems of sample selection and correlated individual heterogeneity in selection and outcome equation simultaneously. We discuss and compare the assumptions under which these estimators produce consistent estimates. We show how they can be extended to take account of not strict exogeneity and/or time constant nonlinear 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 females. The parameter we seek to identify is the effect of actual labour market experience on wages. The problems that arise in this application are nonrandom selection, and unobserved individual specific effects, possibly correlated with the regressors. In addition, actual experience is predetermined, and the experience measure is likely to suffer from (time-constant) measurement error, due to reliance on biographical information before the start of the panel.

We show estimates from different benchmark estimators. OLS/RE do not control for correlated individual heterogeneity, sample selection and measurement error. Standard estimators that difference out the individual effects do not take care of sample selection acting through the idiosyncratic error in the regression of interest, of not strict exogeneity of experience in the

[31] Another way of testing for sample selection is using a Hausman-type test, comparing estimators in columns (11) and (12) with (4) and (6), respectively.

difference equation, or measurement error. Instrumental variable techniques applied to the standard difference estimators may correct for the not strict exogeneity and measurement error problems but not for sample selection. Tests reject random effects specifications (against fixed effects specifications), and exogeneity and/or absence of measurement error of the experience variables in the difference specification.

When we turn to estimation results which take account of a selection process that operates both through the idiosyncratic error and the individual effect in the equation of interest, the estimator by Kyriazidou (1997) avoids specifying the functional form of the sample selection effects, 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 nonsystematic 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 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%. 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. The corresponding IV estimates are smaller, but not precisely estimated.

With Wooldridge's (1995) estimator, the null hypothesis of no correlated individual effects is rejected for all specifications. Conditional on individual effects, the null hypothesis of no sample selection cannot be rejected. Using this estimator, we reject the specification which does not allow for predetermined regressors and measurement error.

Rochina-Barrachina´s estimators indicate that there is nonrandom sample selection, but the joint hypothesis of strict exogeneity of the experience variable and the absence of measurement error can not be rejected, conditional on taking care of correlated heterogeneity and sample selection.

The most general estimator using Wooldridge's (1995) method implies an increase in wages by 1.8% for 1 year of labour market experience, evaluated at 14 years of experience. According to this estimator, the return to experience decreases from 3.1% for the first year to 2.2% after 10 years to 1.2% after 20 years (see Table A.2). Estimates of Rochina-Barrachina's (1999) for all specifications are lower. They range, on average, from 2.3% after the first year to 1.6% after 10 years to 0.8% after 20 years.

The finding that conditional on individual effects, the null hypothesis of no sample selection cannot be rejected with Wooldridge´s (1995) estimator, but is clearly rejected with Rochina-Barrachina's (1999) estimators, may be due to different parametric assumptions imposed by the two methods. The test on selection bias for Wooldridge's estimator requires assumption *W3* to be maintained under the null, which is stronger than what is required for difference fixed effects estimators (for instance the RB estimators).[32]

What are our conclusions? We illustrate that there are several easily implementable estimators available that offer solutions to estimation problems facing the researcher in many typical economic applications. However, our application also illustrates considerable sensitivity of estimates to the particular estimator that is used. As we discuss in detail, this may be due to

[32]In selection correction estimators based on differences (in contrast to level equations) the individual effect in the main equation is differenced out, and we do not have to impose any particular parametric shape to allow for correlated individual heterogeneity and/or selection acting through the individual effect.

different parametric assumptions imposed by the different estimation methods, as well as lack of identifying variation in the data. This suggests caution when interpreting any set of estimates in isolation.

ACKNOWLEDGEMENTS

We are grateful to Richard Blundell, Francois Laisney, Arthur van Soest, Frank Windmeijer and three anonymous referees for useful comments and suggestions. Financial support from the Spanish foundation “Fundación Ramón Areces”, the Spanish Ministry of Science and Technology (Project number SEJ2005-05966 and SEJ2005-08783-C04-01) and the Generalitat Valenciana (Project number GV05/183) to María E. Rochina-Barrachina is gratefully acknowledged. All the GAUSS programmes used for estimation in this paper as well as asymptotic distributions and variances for the estimators are available on http://www.econ.ucl.ac.uk/displayPage.php?page=/downloads/dustmann/index.php. The usual disclaimer applies.

REFERENCES

Altonji, J. G. and R. A. Shakotko (1987). Do wages rise with job seniority?. *Review of Economic Studies 54*, 437–59.

Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press, Advanced Texts in Econometrics.

Blau, F. and L. M. Kahn (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. *Journal of Labor Economics 15*, 1–42.

Blundell, R. and T. MaCurdy (1999). Labour supply: A review of alternative approaches. In:O. Ashenfelter and D. Card (Eds), *Handbook of Labor Economics*, Vol. III, North-Holland: Amsterdam.

Browning, M., Deaton, A., and M. Irish (1985). a profitable approach to labour supply and commodity demand over the life cycle, *Econometrica 53*, 503–43.

Cain, G. G. (1986). The economics analysis of labor market discrimination: A survey. In: *Handbook of Labor Economics*, Vol.1,O. Ashenfelter and R. Layards (Eds), Ch. 13, Amsterdam: Elsevier, 693–785.

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

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies 47*, 225–38.

Chamberlain, G. (1984). Panel data. In *Handbook of Econometrics*, Z. Griliches, M. D. Intriligator (Eds), Vol. II, Ch. 22Amsterdam: North-Holland, 1247–317.

Charlier, E., B. Melenberg and A. Van Soest (1997). An analysis of housing expenditure using semiparametric models and panel data. *CentER Discussion Paper*, no. 9714, Tilburg University, The Netherlands.

Dustmann, C. and C. Meghir (2005). Wages, Experience and Seniority. *Review of Economic Studies 72*, 77–108.

England, P., G. Farkas, B. Stanek Kilbourne and T. Dou (1988). Explaining Occupational Sex Segregation and Wages: Findings from a Model with Fixed Effects. *American Sociological Review 53*(4, 544–58.

Fitzenberger, B. and G. Wunderlich (2002). Gender Wage Differences in Germany: A Cohort Analysis. *German Economic Review 3*, 379–414.

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

Honore, B. and A. Lewbel (2002). Semiparametric Binary Panle Data Models without Strictly Exogenous Regressors. *Econometrica 70*, 2053–64.

Hsiao, C. (1986). *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Kim, M. K. and S. W. Polachek (1991). Panel estimates of male-female earnings functions. *Journal of Human Resources 29*, 406–28.

Kyriazidou, E. (1997). Estimation of a panel data sample selection model. *Econometrica 65*, 1335–64.

Kyriazidou, E. (2001). Estimation of dynamic panel data sample selection models. *Review of Economic Studies 68*, 543–72.

Laisney, F. and M. Lechner (2003). Almost consistent estimation of panel probit models with 'small' fixed effects. *Econometric Reviews 22(1)*, 1–28.

Lee, M. J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. New York: Springer.

Lee, M. J. (2002). *Panel Data Econometrics – Methods of Moments and limited dependent variables*. San Diego: Academic Press.

Lewbel, A. (2003). *Endogenous Selection or Treatment Model Estimation*. Mimeo, Boston College, May 2003.

MACurdy, T. E. (1981). An empirical model of labor supply in a life cycle setting. *Journal of Political Economy 89*, 1059–85.

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

Nijman, T. and M. Verbeek (1992). Nonresponse in panel data: the impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics 7*, 243–57.

Polachek, S. W. and M. K. Kim (1994). Panel Estimates of the Gender Earnings Gap. *Journal of Econometrics 61*, 23–42.

Powell, J. L. (1994). Estimation of semiparametric models. *Handbook of Econometrics 4*, 2444–521.

Rochina-Barrachina, M. E. (1999). A new estimator for panel data sample selection models. *Annales d'Économie et de Statistique 55/56*, 153–81.

Tallis, G. M. (1961). The moment generating function of the truncated multinormal distribution. *Journal of the Royal Statistical Society 23*, Series b, 223–9.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics 68*, 115–32.

Wooldridge, J. M. (2002). *Econometric Analysis of cross Section and Panel Data*. Cambridge, Massachusets: MIT Press.

Wagner, G., R. Burkhauser and F. Behringer (1993). The English language public use file of the German Socio-Economic Panel. *Journal of Human Resources 27*, 429–33.

Zabel, J. E.(1992). Estimating fixed effects and random effects with selectivity. *Economics Letters 40*, 269–72.

APPENDIX A: TABLES

Table A.1. Estimates for the wage equation[a]

Variable	(1) OLS	(2) RE	(3) FE	(4) DE (OLS)	(5) DE (IV)	(6) DE (GMM)	(7)[b] W (MD)	(8)[c] W (MD) $(Ex\,\hat{p})$	(9)[d] K	(10)[d] K (IV)	(11)[e] RB	(12)[e] RB (GMM)	(13)[b] Heckman (2-steps)	(14)[b] Heckman (full MLE)
CST	0.9990*	0.9805*					1.1111*	1.0162*					1.0820*	1.0965*
	(0.0310)	(0.0672)					(0.0724)	(0.0804)					(0.0553)	(0.0391)
D85	0.0047	0.0112	0.0056	0.0052	−0.0062	−0.0041	0.0247	0.0482*	0.0052	0.0052	0.0275	−0.0088	0.0075	0.0078
	(0.0204)	(0.0131)	(0.0139)	(0.0068)	(0.0074)	(0.0029)	(0.0250)	(0.0236)	(0.0068)	(0.0068)	(0.0203)	(0.0087)	(0.0304)	(0.0219)
D86	0.0450*	0.0433*	0.0308	0.0291*	0.0054	0.0168*	0.0466	0.0556*	0.0291*	0.0291*	0.0711*	0.0327*	0.0436	0.0430*
	(0.0206)	(0.0132)	(0.0163)	(0.0095)	(0.0115)	(0.0038)	(0.0295)	(0.0264)	(0.0095)	(0.0095)	(0.0223)	(0.0080)	(0.0302)	(0.0218)
D87	0.0773*	0.0774*	0.0588*	0.0597*	0.0238	0.0332*	0.0876*	0.0920*	0.0597*	0.0597*	0.1001*	0.0835*	0.0763*	0.0762*
	(0.0205)	(0.0136)	(0.0199)	(0.0125)	(0.0158)	(0.0046)	(0.0332)	(0.0257)	(0.0125)	(0.0125)	(0.0252)	(0.0108)	(0.0308)	(0.0220)
D88	0.0826*	0.0749*	0.0492*	0.0602*	0.0131	0.0317*	0.1048*	0.1212*	0.0602*	0.0602*	0.1296*	0.0916*	0.0835*	0.0837*
	(0.0213)	(0.0141)	(0.0240)	(0.0159)	(0.0205)	(0.0056)	(0.0409)	(0.0350)	(0.0159)	(0.0159)	(0.0303)	(0.0121)	(0.0313)	(0.0222)
D89	0.1051*	0.1009*	0.0614*	0.0715*	0.0131	0.0341*	0.1128*	0.1142*	0.0715*	0.0715*	0.1635*	0.1165*	0.1043*	0.1042*
	(0.0205)	(0.0144)	(0.0284)	(0.0194)	(0.0254)	(0.0066)	(0.0468)	(0.0347)	(0.0194)	(0.0194)	(0.0358)	(0.0137)	(0.0311)	(0.0221)
D90	0.1399*	0.1337*	0.0941*	0.1048*	0.0355	0.0568*	0.1394*	0.1466*	0.1048*	0.1048*	0.2043*	0.1679*	0.1390*	0.1389*
	(0.0209)	(0.0149)	(0.0330)	(0.0230)	(0.0302)	(0.0077)	(0.0543)	(0.0402)	(0.0230)	(0.0230)	(0.0393)	(0.0149)	(0.0312)	(0.0222)
D91	0.1453*	0.1592*	0.1142*	0.1254*	0.0454	0.0622*	0.1452*	0.1421*	0.1254*	0.1254*	0.2126*	0.1843*	0.1444*	0.1444*
	(0.0213)	(0.0154)	(0.0378)	(0.0268)	(0.0352)	(0.0089)	(0.0582)	(0.0378)	(0.0268)	(0.0268)	(0.0449)	(0.0164)	(0.0308)	(0.0221)
D92	0.1684*	0.1794*	0.1274*	0.1434*	0.0523	0.0766*	0.1909*	0.1605*	0.1434*	0.1434*	0.2342*	0.1987*	0.1656*	0.1651*
	(0.0213)	(0.0159)	(0.0426)	(0.0304)	(0.0403)	(0.0104)	(0.0660)	(0.0403)	(0.0304)	(0.0304)	(0.0508)	(0.0174)	(0.0309)	(0.0221)

Table A.1. (Continued).

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)[b]	(8)[c]	(9)[d]	(10)[d]	(11)[e]	(12)[e]	(13)[b]	(14)[b]
	OLS	RE	FE	DE	DE	DE	W	W	K	K	RB	RB	Heckman	Heckman
				(OLS)	(IV)	(GMM)	(MD)	(MD) $(Ex\hat{p})$		(IV)		(GMM)	(2-steps)	(full MLE)
D93	0.1683*	0.1839*	0.1258*	0.1439*	0.0422	0.0706*	0.1919*	0.1991*	0.1439*	0.1439*	0.2495*	0.2688*	0.1657*	0.1653*
	(0.0221)	(0.0166)	(0.0475)	(0.0342)	(0.0453)	(0.0109)	(0.0719)	(0.0430)	(0.0342)	(0.0342)	(0.0556)	(0.0195)	(0.0313)	(0.0222)
D94	0.1724*	0.1917*	0.1276*	0.1461*	0.0335	0.0628*	0.2227*	0.2073*	0.1461*	0.1461*	0.2474*	0.2769*	0.1707*	0.1703*
	(0.0215)	(0.0174)	(0.0525)	(0.0380)	(0.0502)	(0.0112)	(0.0790)	(0.0455)	(0.0380)	(0.0380)	(0.0602)	(0.0218)	(0.0316)	(0.0223)
D95	0.2159*	0.2119*	0.1398*	0.1594*	0.0367	0.0668*	0.2591*	0.2519*	0.1594*	0.1594*	0.2736*	0.3556*	0.2160*	0.2160*
	(0.0225)	(0.0182)	(0.0572)	(0.0415)	(0.0549)	(0.0131)	(0.0856)	(0.0512)	(0.0415)	(0.0415)	(0.0659)	(0.0238)	(0.0326)	(0.0226)
ED	0.1133*	0.1125*					0.1065*	0.1086*					0.1102*	0.1097*
	(0.0020)	(0.0054)					(0.0042)	(0.0043)					(0.0029)	(0.0023)
EXP	**0.0309***	**0.0274***	**0.0349***	**0.0324***	**0.0522***	**0.0473***	**0.0230***	**0.0320***	**0.0525***	**0.0157**	**0.0244***	**0.0229***	**0.0303***	**0.0301***
	(0.0019)	(0.0023)	(0.0062)	(0.0042)	(0.0058)	(0.0017)	(0.0090)	(0.0060)	(0.0222)	(0.1935)	(0.0060)	(0.0021)	(0.0034)	(0.0019)
EXP2	**−0.0058***	**−0.0045***	**−0.0045***	**−0.0044***	**−0.0065***	**−0.0060***	**−0.0029***	**−0.0049***	**−0.0041**	**−0.0014**	**−0.0041***	**−0.0047***	**−0.0055***	**−0.0055***
	(0.0005)	(0.0005)	(0.0005)	(0.0002)	(0.0003)	(0.0001)	(0.0009)	(0.0012)	(0.0050)	(0.0470)	(0.0005)	(0.0002)	(0.0007)	(0.0005)
$\partial w/\partial Exp$**(14 years)**	**0.0148***	**0.0147***	**0.0223***	**0.0200***	**0.0340***	**0.0305***	**0.0148***	**0.0182***	**0.0409***	**0.0116**	**0.0129***	**0.0097***	**0.0148***	**0.0148***
	(0.0007)	(0.0013)	(0.0056)	(0.0039)	(0.0054)	(0.0014)	(0.0077)	(0.0038)	(0.0105)	(0.0637)	(0.0054)	(0.0017)	(0.0014)	(0.0008)
Av. ret. T. dummies	0.1204*	0.1243*	0.0850*	0.0953*	0.0268	0.0461*	0.1387*	0.1399*	0.0953*	0.0953*	0.1739*	0.1607*	0.1197*	0.1195*
	(0.0150)	(0.0119)	(0.0336)	(0.0228)	(0.0301)	(0.0075)	(0.0492)	(0.0306)	(0.0228)	(0.0228)	(0.0380)	(0.0137)	(0.0233)	(0.0165)

Notes: [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 use of pre-estimated time dummies coefficients obtained from the simple difference estimator in (4).[e]Standard errors corrected for the first stage maximum likelihood bivariate probit estimates. *Statistically different from zero at the 5% significance level.

Table A.2. Estimated rates of return for work experience $(\partial w/\partial Exp)^{(a)}$

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

Notes: [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 use of pre-estimated time dummies coefficients obtained from the simple difference estimator in (4). [e]Standard errors corrected for the first stage maximum likelihood bivariate probit estimates. *Statistically different from zero at the 5% significance level.

APPENDIX B: NOTES TO TABLE 5

[1]The test statistic for H_0: $\beta_{FE} = \beta_{RE}$ is $(\beta_{FE} - \beta_{RE})' \cdot [Vce(\beta_{FE}) - Vce(\beta_{RE})]^{-1} \cdot (\beta_{FE} - \beta_{RE}) \Rightarrow \chi^2{}_{13}$.

[2]The test statistics for H_0: $\beta_{DE(IV)} = \beta_{DE(OLS)}$ or $\beta_{DE(GMM)} = \beta_{DE(OLS)}$ are $(\beta_{DE(IV)} - \beta_{DE(OLS)})' \cdot [Vce(\beta_{DE(IV)} - \beta_{DE(OLS)})]^{-1} \cdot (\beta_{DE(IV)} - \beta_{DE(OLS)}) \Rightarrow \chi^2_{14}$, or $(\beta_{DE(GMM)} - \beta_{DE(OLS)})' \cdot [Vce(\beta_{DE(GMM)} - \beta_{DE(OLS)})]^{-1} \cdot (\beta_{DE(GMM)} - \beta_{DE(OLS)}) \Rightarrow \chi^2_{14}$, respectively.

[3]For Wooldridge's estimators, Wald tests for the linear hypothesis H_0: $\ell_{84} = 0, \ell_{85} = 0, \ldots, \ell_{95} = 0$. The test statistics for H_0 are $(\Re\ \beta_{W(MD)})' \cdot [\ Vce(\Re\ \beta_{W(MD)})]^{-1} \cdot (\Re\ \beta_{W(MD)}) \Rightarrow \chi^2_{12}$ and $(\Re\beta_{W(MD)_{(Ex\hat{p})}})\prime \cdot [Vce(\Re\beta_{W(MD)_{(Ex\hat{p})}})]^{-1} \cdot (\Re\beta_{W(MD)_{(Ex\hat{p})}}) \Rightarrow \chi^2_{12}$, respectively, where $\Re$ is the matrix of zeros and ones that allows selecting the 12 lambda terms coefficients to be able to test for the previous linear hypothesis. For Rochina-Barrachina's estimators, Wald tests for the linear hypothesis H_0: $\ell_{84,85} = 0, \ell_{85,84} = 0, \ldots, \ell_{94,95} = 0, \ell_{95,94} = 0$. The test statistics for H_0 are $(\mathbb{R}\beta_{RB})' \cdot [\ Vce(\mathbb{R}\beta_{RB})]^{-1} \cdot (\mathbb{R}\beta_{RB}) \Rightarrow \chi^2_{66*2}$ and $(\mathbb{R}\beta_{RB(GMM)})' \cdot [\ Vce(\mathbb{R}\beta_{RB(GMM)})]^{-1} \cdot (\mathbb{R}\beta_{RB(GMM)}) \Rightarrow \chi^2_{66*2}$, respectively, where $\mathbb{R}$ is the matrix of zeros and ones that allows selecting the 132 lambda terms coefficients to be able to test for the previous linear hypothesis.

[4]Wald tests for the linear hypothesis H_0: $\psi_{Exp} = 0, \theta_{Exp^2} = 0$. The test statistics for H_0 are $(\bar{\Re}\beta_{W(MD)})' \cdot [Vce(\bar{\Re}\beta_{W(MD)})]^{-1} \cdot (\bar{\Re}\beta_{W(MD)}) \Rightarrow \chi^2_2$ and $(\bar{\Re}\beta_{W(MD)_{(Ex\hat{p})}})' \cdot [Vce(\bar{\Re}\beta_{W(MD)_{(Ex\hat{p})}})]^{-1} \cdot (\bar{\Re}\beta_{W(MD)_{(Ex\hat{p})}}) \Rightarrow \chi^2_2$, respectively, where $\bar{\Re}$ is the matrix of zeros and ones that allows selecting the two coefficients for average experience and average squared experience to be able to test for the previous linear hypothesis.

[5]The test statistic for H_0: $R\beta_K = R\beta_{DE(OLS)}$ is $(R\beta_K - R\beta_{DE(OLS)})' \cdot [Vce(R\beta_K - R\beta_{DE(OLS)})]^{-1} \cdot (R\beta_K - R\beta_{DE(OLS)}) \Rightarrow \chi^2_2$, where R is the matrix of zeros and ones that allows selecting the two coefficients for experience and squared experience to be able to test for the previous hypothesis.

[6]The test statistics for H_0: $\beta_{W(MD)_{(Ex\hat{p})}} = \beta_{W(MD)}$ or $\bar{\mathbb{R}}\beta_{RB(GMM)} = \bar{\mathbb{R}}\beta_{RB}$ are $(\beta_{MD_{(Ex\hat{p})}} - \beta_{MD})' \cdot [Vce(\beta_{MD_{(Ex\hat{p})}} - \beta_{MD})]^{-1} \cdot (\beta_{MD_{(Ex\hat{p})}} - \beta_{MD}) \Rightarrow \chi^2_{29}$ or $(\bar{\mathbb{R}}\beta_{RB(GMM)} - \bar{\mathbb{R}}\beta_{RB})' \cdot [Vce(\bar{\mathbb{R}}\beta_{RB(GMM)} - \bar{\mathbb{R}}\beta_{RB})]^{-1} \cdot (\bar{\mathbb{R}}\beta_{RB(GMM)} - \bar{\mathbb{R}}\beta_{RB}) \Rightarrow \chi^2_2$, respectively, where $\bar{\mathbb{R}}$ is the matrix of zeros and ones that allows selecting the two coefficients for experience and squared experience to be able to test for the last hypothesis.