On the estimation and testing of wage equations using GMM

Samuel G. Berlinski[1]
Nuffield College, Oxford University
Email:samuel.berlinski@nuf.ox.ac.uk

First version: 18 April 2000. This version: 3 August 2000.

[1] I want to thank Steve Bond for his encouragment, extensive discussion, and suggestions. The usual disclaimers apply.

Abstract

We apply GMM techniques to the estimation and testing of a wage equation using data from the British Household Panel Survey. We cannot reject the strict exogeneity of size, union and industry with respect to the time varying idiosyncratic shock. We find compelling evidence of correlation between the unobserved time invariant individual effect, η_i, and size and industry. However, there is no evidence of correlation between η_i and union. None of our diagnostic tests' suggest a strong correlation between schooling and η_i. In fact, OLS levels estimates of the schooling coefficient are downward biased because of inconsistency in other estimates.

JEL classification: C23; C52; J31

Keywords: Panel data; Specification testing; GMM; Wage equations; Wage differentials.

1 Introduction

The estimation of wage differentials has always been a matter of great interest for economists. For example, we have tried to estimate the effects on wages of an additional year of schooling[1], of union status[2], of industry affiliation[3], and of employer size[4] among many other things. There have always been worries in this literature that the omission of individual characteristics correlated with the variables of interest and wages may bias the cross-sectional estimates obtained by, for example, ordinary least squares (OLS). Researchers have attempted a different range of econometric techniques to deal with this problem, from relying on panel data in order to exploit variation over time within cross-sectional units[5] using so called fixed-effect transformations, to the use of quasi-natural experiments that exploit cross-sectional variation in the variable of interest that is uncorrelated with the omitted variables[6].

This paper is related to the panel data literature that uses transformations of the data to control for omitted individual characteristics. Although the fixed-effect transformation solves the problem posed by the omission of time invariant characteristics the solution may come at some costs. First, measurement error issues, that tend to be negligible when only information in levels is used, are exacerbated by fixed-effect transformations severely biasing the estimated coefficients of the time varying characteristics of interest towards zero. Second, by transforming the data to control for omitted individual characteristics, we cannot identify the coefficients for time invariant characteristics (e.g., schooling). The solution for both problems has been to use instrumental variables techniques, relying on set a of internal instruments, to obtain consistent estimates of the parameters of interest[7]. Unfortunately, the conditions under which these instruments are valid are often neither clear nor thoroughly tested.

[1] For example, Mincer (1974), Grilliches (1977), Chowdhury and Nickell (1985), and Angrist and Krueger (1991).

[2] For example, Duncan and Leigh (1980), Freeman (1984), Chowdhury and Nickell (1985), Robinson (1989), and Jakubson (1991).

[3] For example, Murphy and Topel (1987), Krueger and Summers (1988), Keane (1993), and Goux and Maurin (1999).

[4] For example, Brown and Meadoff (1989).

[5] For example, Chowdhury and Nickell (1985), Murphy and Topel (1987), Krueger and Summers (1988), Brown and Meadoff (1989), Jakubson (1991), Hughes (1998) and Goux and Maurin (1999).

[6] For example, Angrist and Krueger (1991).

[7] For example, Nickell and Chowdhury (1985), Cornwell and Rupert (1988), Murphy and Topel (1987), and Goux and Marin (1999).

In order to deal with these issues we use a family of generalized method of moments (GMM) estimators for panel data, with a large cross-sectional dimension and fixed (or short) time-series dimension, that have been proposed by Arellano and Bond (1991) and Arellano and Bover (1995). These estimators have been widely applied to company data. They are not only suitable to deal with omitted time invariant characteristics and measurement error, but also with the fact that some of the time varying variables of interest may only be predetermined with respect to the time varying error component. This problem has received relatively little attention in the estimation of wage equations[8],[9]. This despite the fact that, in the absence of strict exogeneity and with the typically short length of available panels, estimators such as random-effects, within-groups or OLS in first-differences are inconsistent.

It is worth pointing out at this stage that this paper is not concerned with the performance of the many different estimators that have been proposed in the literature to deal with some or all of the problems we have pointed out here (e.g., Cornwell and Rupert (1988) or Ziliak (1997)). The purpose of the paper is to present a framework where the conditions under which different estimators are consistent and/or can increase efficiency are clear. Moreover, we like to stress that many of these conditions are testable so, in principle, it may be reasonable to test rather than assume these conditions as we often do when we estimate wage differentials using panel data.

The outlay for the rest of the paper is the following. Section 2 is divided in two. First, in 2.1, we present different scenarios for the specification of the wage equation and highlight the conditions under which different estimators are consistent. Second, in 2.2, we define the GMM estimators of Arellano and Bond (1991) and Arellano and Bover (1995) and stress a battery of specification tests that are available in this context. Section 3 applies GMM estimation techniques and specification tests to the estimation of a wage equation using the eight available waves from the British Household Panel Survey (BHPS). Section 4 discusses the results further and

[8]The exception has been the seminal work from Chamberlain (1982, 1984), and the papers from Angrist and Newey (1991) and Jakubson (1991) that followed, where the appropiateness of the fixed-effects model is investigated by testing the overidentifying restrictions it places on the reduced form.

[9]In the area where people have been most concerned about this problem is in the estimation of rational expectations models using either time-series or panel data. See, for example, in the time-series literature Hayashi and Sims (1983) and in the panel data literature the paper by Keane and Runkle (1992) and comments following this paper.

concludes.

2 Model specification and method of estimation

2.1 Model specification

Consider the following wage equation

$$\begin{aligned} w_{it} &= \beta' x_{it} + \gamma' f_i + u_{it}, \quad t = 1, ..., T, \quad i = 1, ..., N, \\ u_{it} &= \eta_i + \lambda_t + \upsilon_{it} \end{aligned} \tag{1}$$

where w_{it} is log hourly wage, x_{it} is a $k \times 1$ vector of time varying explanatory variables that will typically include experience, experience squared, union status, industry affiliation, establishment size, marital status, occupation and location, and f_i is a $g \times 1$ vector of time invariant explanatory variables that will typically include years of schooling, gender and ethnic background. The number of periods T is fixed and the number of individuals N is large.

The disturbances have a two-way error component structure, where η_i is an unobservable (for the econometrician) individual effect and λ_t is a time specific effect. The η_i can be interpreted as a time invariant worker characteristic (e.g., school quality or 'look') that is observable for the employer, equally rewarded in all jobs but unobservable for the econometrician (i.e., a variable omitted from the wage equation). Furthermore, we assume that the observations are independently distributed over the cross-section and that, $E(\eta_i) = 0$, $E(\upsilon_{it}) = 0$, $E(\upsilon_{it}\eta_i) = 0$ for $i = 1, ..., N$ and $t = 1, ..., T$, and $E(\upsilon_{it}\upsilon_{is}) = 0$ for $i = 1, ..., N$ and for any t different than s. Notice that since T is fixed and there is independence in the cross-sectional dimension for the υ_{it}s we can control for the time-specific effects by including year dummies in the regression function.

The purpose of this section is to analyze which are the appropriate estimators for β and γ under different sets of assumptions regarding the stochastic behavior of u_{it}, x_{it} and f_i. First, consider the case where the x_{it} and f_i are strictly exogenous with respect to υ_{is},

$$E(\upsilon_{it}/x_{i1}, ..., x_{iT},\ f_i, \eta_i, \lambda_t) = 0, \tag{2}$$

there is mean independence of η_i given $x_{i1}, ..., x_{iT}$, f_i, and λ_t,

$$E(\eta_i/x_{i1}, ..., x_{iT},\ f_i, \lambda_t) = 0, \tag{3}$$

and homoskedasticity across both individuals and time (i.e., $E(\upsilon_{it}^2) = \sigma_\upsilon^2$ and $E(\eta_i^2) = \sigma_\eta^2$). Under these conditions, the standard generalized least squares (GLS) estimator proposed by Balestra and Nerlove (1966) will provide consistent and efficient estimates for β and γ[10].

What happens with GLS estimates of (1) when either (2) or (3) or both do not hold? Let me start by considering the implications of the assumption of strict exogeneity of time varying variables with respect to the time varying idiosyncratic shock. If that condition were to hold it will rule out both feedback effects from the shocks, υ_{is}, to the x_{it}s and measurement error in the x_{it}s. There are reasonable scenarios where the absence of feedback seems unlikely. Consider two possible examples. First, suppose that the firm where individual i is working has lost an important client in period t this will affect i's earnings at t and may trigger a job change (and hence industry, union status, etc.) at $t + s$ where $s \geq 1$. Second, suppose that an individual i is affected by some illness in period t, this illness will affect i's earnings at t and even if i makes a full recovery by $t + 1$[11] he/she may change jobs at some point in the future, say, because he/she may have lost out on a promotion opportunity to another worker.

In the absence of measurement error and assuming that the x_{it} are only predetermined with respect to υ_{is},

$$E(\upsilon_{it}/x_{i1}, ..., x_{it},\ f_i, \eta_i, \lambda_t) = 0, \tag{4}$$

the GLS estimation of the model (1), (4) and (3) is inconsistent when the time dimension is fixed. The reason for this is the following, GLS is equivalent to running OLS to a transformed model where the transformation for x_{it}, say, is $\widetilde{x}_{it} = x_{it} - \theta\overline{x}_i$ with $\theta = 1 - \sigma_\upsilon/\sqrt{\sigma_\upsilon^2 + T\sigma_\eta^2}$, $0 < \theta < 1$, and $\overline{x}_i = T^{-1}\sum_{t=1}^{T} x_{it}$. The inconsistency is due to the asymptotic correlation between $(x_{it} - \theta\overline{x}_i)$ and $(\upsilon_{it} - \theta\overline{\upsilon}_i)$. Though x_{it}

[10] OLS is under these assumptions consistent but inefficient because of the existence of autocorrelation.

[11] Obviously this person might not make a full recovery by $t+1$ which will imply that the static model is implausible and some sort of persistence should be introduced (e.g., the υ_{it} can be an AR(1) process if the effect decline exponentially but it never quite disappears or a MA(1) process if the effect has a short finite memory). Fortunately, this is a testable implication.

and υ_{it} are uncorrelated, their respective individual means are correlated with each other, with υ_{it} and with x_{it}, and the sum of those three covariances does not vanish as N tends to infinity. A possible solution for estimating consistent parameters for the model defined by (1), (3), and (4) is to ignore η_i and to estimate the model in levels by simply applying OLS[12],[13]. Clearly, if there are in fact individual effects, there is going to be autocorrelation so a consistent estimate of the covariance matrix is required in order to make statical inference about the parameters of interest.

In a typical measurement error framework with uncorrelated measurement errors and with the 'true' x_{it} strictly exogenous with respect to υ_{is}

$$E(\upsilon_{it}/x_{i1},...,x_{it-1},x_{it+1},...,x_{iT},\ f_i,\eta_i,\lambda_t)=0. \tag{5}$$

Now, both GLS and OLS produce inconsistent estimates of the parameters of interest for the model defined by (1), (3) and (5). However, provided that the 'true' x_{it}s are serially correlated, an instrumental variable estimator can provide consistent estimates for the parameters of interest by exploiting the fact that $x_{i1},...,x_{it-1},x_{it+1},...,x_{iT}$ are uncorrelated to υ_{it}. If the 'true' x_{it} are only predetermined with respect to υ_{is}, the potential set of instruments available is reduced to $x_{i1},...,x_{it-1}$.

What happens if the assumption of mean independence of the individual effects with respect to, at least, some of the x_{it} and the f_i does not hold? This is to say,

$$E(\eta_i/x_{i1},...,x_{iT},\ f_i,\lambda_t)\neq 0. \tag{6}$$

In this case, both GLS and OLS levels estimators of the model (1), (2) and (6) are inconsistent. Under strict exogeneity of the x_{it} with respect to the υ_{is}, OLS regression of a transformed model that sweeps out the individual effect η_i (e.g., first-differences, deviations around individual means or orthogonal deviations) will allow us to obtain a consistent estimate of β. Moreover, provided that either the f_i are uncorrelated with η_i or that there are enough instruments for those f_i that are correlated with the individual effect, γ can also be identified (see, Hausman and Taylor (1981), Amemiya and MaCurdy (1986), Breusch, Mizon, and Schmidt (1989) and, Arellano and Bover (1995)). This may allow us to identify the coefficient of non

[12] Of course, when $\theta=0$ the GLS estimate reduces to OLS levels.

[13] If lagged dependent variables are included in model (1), OLS will produce inconsistent parameter estimates (see, Hsiao (1986)). Instrumental variables will produce consistent parameter estimates provided that we can find instruments that are uncorrelated with η_i that can be used to instrument the lagged dependent variables.

time varying aspects (e.g., schooling) and the linear effects of time varying aspects which increment identically with time (e.g., experience).

It is worth noticing in this context that, even when all x_{it} and f_i are suspected of being correlated with the individual effect, there may be instruments to identify the coefficient of time invariant variables. Bhargava and Sargan (1983) and Breusch, Mizon and Schmidt (1989) exploit the instruments that arise if the correlation between x_{it} and η_i is assumed constant[14]. A (stronger) conditional expectation version of this assumption yields, for example[15],

$$E(\eta_i/\triangle x_{i2}, ..., \triangle x_{iT}, \lambda_t) = 0; \tag{7}$$

provided that the $\triangle x_{i2}, ..., \triangle x_{iT}$ are correlated with the f_i, consistent estimates of γ can also be obtained.

In practice, however, finding internal instruments that are both uncorrelated with the unobserved individual effect and correlated with the observed f_i may be difficult. As Chowdhury and Nickell (1985) point out, in relation to their attempts to instrument the schooling variable: "In all our models where the instruments pass the "exogeneity" test, the coefficient on schooling is weak and poorly determined ... Basically, we require better valid instruments than appear to be available within our data set in order to pin down the schooling effect with any precision" (p. 62).

However, the identification of the coefficients of time invariant variables and the linear effects of time varying aspects which increment identically with time is not the only rational for the use of the additional moment conditions implied by (7). Even if we were only interested in identifying the effect on wages of time varying variables, the use of the additional level moment conditions may increase the efficiency with which time varying coefficients are estimated, especially in short panels (see, Arellano (1993) and Blundell and Bond (1998)).

Although the fixed-effect transformation solves the problem posed by the unobservability of η_i for the estimation of the coefficients of time varying characteristics, the solution comes at some cost. Measurement error issues that may have been

[14] As noted in Arellano and Bover (1995).

[15] Breush, Mizon and Pagan (1989) use

$$E(\eta_i/x_{i1} - \overline{x}_i, ..., x_{iT} - \overline{x}_i) = 0.$$

otherwise negligible for estimation in levels are exacerbated by fixed-effect transformations, severely biasing the estimates of β towards zero[16]. Moreover, even in the absence of measurement error, if the x_{it} are only predetermined with respect to the v_{is} (see condition (4)) after applying any fixed-effect transformation to the data, OLS estimation of β is inconsistent. For example, an OLS estimation of a model in first-differences is inconsistent[17] because

$$E(v_{it} - v_{it-1}/x_{it}) \neq 0.$$

Griliches and Hausman (1986) note that an instrumental variables estimator can be used to obtain consistent estimates of β. For example, in a measurement error framework with uncorrelated measurement errors, where the 'true' variables are predetermined, and assuming that the 'true' values of $x_{i1}, ..., x_{iT}$ are serially correlated the following relation can be exploited

$$E(v_{it} - v_{it-1}/x_{i1}, ..., x_{it-2},\ f_i, \lambda_t) = 0. \tag{8}$$

In fact, this idea can be extended to any fixed-effect transformation that eliminates the individual effect from the transformed error term, without at the same time introducing all lagged values of v_{it} into the transformed error term. Orthogonal deviations (see, Arellano and Bover (1995)) is another possible example of this sort of transformation. The fact that these transformations do not introduce all lagged values of the disturbances into the transformed error term allow us to use suitably lagged time varying variables as instruments for the transformed model.

It worth noticing that with respect to the situation described by (8), in the absence of measurement error or when the 'true' variables are strictly exogenous, additional instruments are available. For example, in the absence of measurement error, (8) becomes

$$E(v_{it} - v_{it-1}/x_{i1}, ..., x_{it-1},\ f_i, \lambda_t) = 0, \tag{9}$$

[16] The extent of the bias from measurement error produced by the different transformations is likely to differ and may provide an important identification tool. See, Griliches and Hausman (1986).

[17] In fact, even if it is consistent, it will in general be inefficient because it does not take into account the first order serial correlation produced by first differencing.

or under uncorrelated random measurement error in the time varying explanatory variables but with the 'true' x_{it}s strictly exogenous, (8) becomes

$$E(\upsilon_{it} - \upsilon_{it-1}/x_{i1}, ..., x_{it-2}, x_{it+1}, ..., x_{iT}, f_i, \lambda_t) = 0. \qquad (10)$$

Even if there is no strict exogeneity of the x_{it}, either because of feedback effects and/or classical measurement error, there are still moment conditions in levels that might help to identify γ or to increase the efficiency with which we estimate β or both. However, the set of internal instruments available in the absence of strict exogeneity is restricted. For example, if condition (7) holds, we have the following moment condition

$$E(u_{it}/\triangle x_{i2}, ..., \triangle x_{it-1}, \lambda_t) = 0$$

when the 'true' x_{it}s are predetermined and there is uncorrelated measurement error.

Finally, a standard procedure in the estimation of wage equations has been to use a Hausman[18] test in order to compare the within-group estimates of the parameters of interest with the more efficient, under the null hypothesis of mean independence (3), GLS estimates. It is worth noticing that, for the test in the standard econometric package to shed some light on the existence of any endogeneity affecting the estimates of the parameters of interest, it should be the case that there is no heteroskedasticity, no autocorrelation and strict exogeneity of the x_{it} with respect to the υ_{is}. When there is heteroskedasticity and/or serial correlation none of the previous estimators are optimal under the null or the alternative hypothesis. Moreover, in the absence of strict exogeneity and with the time dimension of the panel short none of them are consistent.

In the next part of this Section, we review some GMM estimators for panel data suitable for dealing with predetermined variables, measurement error and absence of mean independence as well as some specification tests that are available in this framework.

2.2 A generalized method of moments estimator and some specification tests

Given our concern with the possible absence of strict exogeneity in the x_{it}s with respect to υ_{is} and the possible correlation between the x_{it}s, f_is, and η_i we will

[18] See, for example, Hausman and Taylor (1981).

estimate the wage equation using an instrumental variable approach in which, the instruments are weighted optimally so as to form a Generalized Method of Moments (GMM) estimator[19].

Given the possible correlation between the x_{it}s, f_is, and the individual effect, η_i, we use a first-differences transformation to estimate β. In such case, the GMM estimator is (omitting time effects[20])

$$\widehat{\beta} = \left[\left(\sum_i \triangle x_i' Z_i^d \right) A_N \left(\sum_i Z_i^{d\prime} \triangle x_i \right) \right]^{-1} \left(\sum_i \triangle x_i' Z_i^d \right) A_N \left(\sum_i Z_i^{d\prime} \triangle w_i \right), \tag{11}$$

where $\triangle x_i$ is a stacked vector of $\triangle x_{it}$ and $\triangle w_i$ is a stacked vector of observations on $\triangle w_{it}$. Conditions (10), (9), and (8) established different sets of instruments available under different scenarios. If there is uncorrelated measurement error, the 'true' x_{it} are predetermined with respect to υ_{is}, and we allow the implicit reduced form to be different in each cross-section the instrument matrix, Z_i^d has the form[21],[22],[23]

$$Z_i^d = \begin{bmatrix} x_{i1} & 0 & 0 & 0 & . & . & 0 & . & 0 \\ 0 & x_{i1} & x_{i2} & 0 & . & . & 0 & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & x_{i1} & . & x_{iT-2} \end{bmatrix}. \tag{12}$$

For any consistent estimate of β the optimal choice of A_N can be calculated as $\left[N^{-1} \sum_i^N \left(Z_i^{d\prime} \triangle \widehat{u}_i \triangle \widehat{u}_i' Z_i^d \right) \right]^{-1}$. Our preliminary consistent estimates of β are obtained using as a weighting matrix $(N^{-1} \sum_i^N Z_i^{d\prime} H_i^d Z_i^d)^{-1}$, where H_i^d is a matrix with twos on the leading diagonal, minus one on the first off-diagonal, and zero elsewhere[24]. It is worth pointing-out that this estimator does not require that we assume homoskedasticity across time or individuals. In fact, all the GMM standard errors

[19] Hansen (1982) and White (1982) showed that improvements in efficiency over two-stages least squares are possible by optimally weighting the distance between the sample and population moments, with the weights being the inverse of the covariance matrix of the sample moments.

[20] If we include time dummies and a constant, as we do in the empirical section, we need to drop experience from $\triangle x_i$.

[21] If the matrix of instruments has this form, then the first observation on the stacked vectors $\triangle x_i$ and $\triangle w_i$ are $\triangle x_{i3}$ and $\triangle w_{i3}$, respectively. Therefore this estimator requires that $T \geq 3$.

[22] In an unbalanced panel, for individuals with incomplete data the rows of Z_i^d corresponding to the missing equations are deleted, and missing values in the remaining rows are replaced by zeros.

[23] In theory, we could have also used the time-invariant variables, f_i, in the instruments' set for each of the cross sections. However, we suspect that they might be only weakly correlated with the instrumented time varying right-hand side variables and therefore we prefer not to include them.

[24] Notice, however, that if the υ_{it}s are not only serially uncorrelated but also homoskedastic, the first-step estimate is asymptotically equivalent to the two-step estimate.

reported in this paper are asymptotically robust to time-series or cross-sectional heteroskedasticity of unknown type.

In the empirical section that follows we will concentrate in the estimation of wage differentials for time varying characteristics such as employer size, union status and industry affiliation. As we have mentioned in the first part of this Section, 2.1, in addition to the more traditionally used moments conditions in first-differences, moment conditions in levels may also be available even when all x_{it} are suspected of being correlated with the individual effect. The use of the additional moment restrictions may increase the efficiency with which time varying coefficients are estimated (see, Arellano (1993) and Blundell and Bond (1998)).

Assuming an uncorrelated measurement error structure, with the 'true' x_{it}s predetermined with respect to v_{is} and if condition (7) holds, a GMM estimator for β and γ similar to (11), but that also uses level moment conditions, can be constructed as follows: instead of $\triangle w_i$ we use now the stacked vector $\triangle w_i^+ =(\triangle w_{i3}, ..., \triangle w_{iT}, w_{i3}, ..., w_{iT})$, instead of $\triangle x_i$ we use now the stacked vector $\triangle x_i^+ =(\triangle x_{i3}\, 0, ..., \triangle x_{iT}\, 0, x_{i3}\, f_i, ..., x_{iT}$ $f_i)$[25] with 0 a $g \times 1$ vector of zeroes, instead of Z_i^d the matrix of instruments is now

$$Z_i^+ = \begin{bmatrix} Z_i^d & 0 \\ 0 & Z_i^l \end{bmatrix} \tag{13}$$

where if we only suspect that the x_{it} are correlated with η_i and restricting the coefficients in the first stage regression for the f_i to have a common coefficient in all cross-sections Z_i^l has the form[26],

$$Z_i^l = \begin{bmatrix} \triangle x_{i2} & 0 & . & . & 0 & f_i \\ 0 & \triangle x_{i3} & . & . & 0 & f_i \\ . & . & . & . & . & f_i \\ . & . & . & . & \triangle x_{iT-1} & f_i \end{bmatrix}.$$

If we suspect that some of the time varying variables are uncorrelated with the individual effect, we can introduce in Z_i^l the *level* of those variables lagged once instead of introducing the first differences of them. Introducing both will be redundant given the moment conditions for the first-differenced equations. Moreover, if any of the time invariant variables is suspected to be correlated with the individual effect

[25] Notice that we can include now the experience variable in the set of time varying right-hand side variables for both the equations in first-differences and in levels even if we include time dummies.

[26] Although there are more moment conditions for equations in levels if we add to the instruments' set $\triangle x_{it-s}$ for $s > 1$, it can be proved that there are redundant once we take into account the moment conditions for the equations in first-differences. See Arellano and Bover (1995).

(e.g., schooling) it could be excluded from the instruments' set[27]. The weighting matrix for the first-step regression is

$$H_i^+ = \begin{bmatrix} H_i^d & 0 \\ 0 & I_i \end{bmatrix}$$

where I_i is a an identity matrix with dimension $T-2$[28].

Arellano and Bond (1991) and Arellano (1993) propose a battery of specification tests that can be applied in this framework. First, as there are plenty of overidentifying restrictions available if $T > 3$ and different assumptions related to the specification of the wage equation imply that different sets of instruments are available, tests of overidentifying restrictions become a very valuable tool in assessing model specification. The validity of the overidentifying restrictions can be tested[29] using a two-step robust Sargan-Hansen test. The expression for this test for the model in first-differences, for example, is

$$SH = N^{-1}\left(\sum_i^N \triangle\widehat{u}_i' Z_i^d\right) A_N \left(\sum_i^N Z_i^{d\prime} \triangle\widehat{u}_i\right).$$

SH is asymptotically distributed Chi-square under the null that the overidentifying restrictions are valid with degrees of freedom equal to the number of overidentifying restrictions. Also, under different sets of assumptions about the model only a nested set of instruments are valid. This makes possible to use a two-step Sargan-Hansen difference test to assess two possible models. For example, the set of instruments available with uncorrelated measurement error and the 'true' x_{it} predetermined with respect to v_{it}, say L_1, is nested within the set of instruments available with predetermined x_{it} but in the absence of measurement error, say L_2, (see, conditions (8) and (9), respectively). Then, with obvious notation, the Sargan-Hansen difference test is $SH_2 - SH_1$. This test is asymptotically distributed Chi-square under the null that the additional moment conditions are valid with $L_2 - L_1$ degrees of freedom. Notice that here we just referred to two-step Sargan-Hansen tests because these are the only tests robust to heteroskedasticity, which is a pervasive feature of wage equations.

[27] The question is, however, how good an instruments' set is Z_i^+ to identify any of the coefficients of the time invariant characteristics excluded from the instruments' set.

[28] In an unbalanced panel, the dimension will be equal to the number of levels equations observed for individual i. In a balanced panel without measurement error the dimension will be $T-2$.

[29] To be more precise this test provides, under the null, a joint test for the overidentifying restrictions and for other aspects of model misspecification.

Second, it is important for the validity of the whole instrumenting procedure in the first-differenced equations to test whether there is serial correlation in the v_{it}s. For this purpose, we employ the one degree of freedom test m_2 that tests for the absence of second order serial correlation in the disturbances of a model in first-differences. The test is distributed $N(0,1)$ under the null of lack of second order serial correlation in the first-differenced disturbances (by construction we will expect first order serial correlation in the disturbances of a model in first-differences). Again, we use a variance-covariance matrix that is robust to individual and time-series heteroskedasticity for this test.

Finally, we can use the Hausman type tests proposed in Arellano (1993) to test for the absence of mean independence and, more generally, for the validity of the instruments' set for the equations in levels. The idea in Arellano (1993) is to "regard correlated effect biases as misspecification due to the exclusion of relevant variables in a standard regression model" (p. 87). Consider a wage equation, (1), where the x_{it} are predetermined with respect to v_{is}, (4). The null hypothesis is mean independence, (3), and the alternative hypothesis is

$$E(\eta_i / x_{i1}, ..., x_{iT}, f_i, \lambda_t) = \alpha_1' x_{it} + \alpha_2' f_i$$

with α_1 and/or α_2 different from zero. Omitting time effects and assuming for simplicity that $\alpha_2 = 0$

$$\begin{aligned} E(y_i / x_{i1}, ..., x_{it}, f_i) &= \beta' x_{it} + \gamma' f_i + E(\eta_i / x_{i1}, ..., x_{it}, f_i) \\ &= \beta' x_{it} + \gamma' f_i + \alpha_1' x_{it}. \end{aligned}$$

Combining equations in first-differences and in levels

$$\begin{bmatrix} \triangle y_{it} \\ y_{it} \end{bmatrix} = \beta' \begin{bmatrix} \triangle x_{it} \\ x_{it} \end{bmatrix} + \gamma' \begin{bmatrix} 0 \\ f_i \end{bmatrix} + \alpha_1 \begin{bmatrix} 0 \\ x_{it} \end{bmatrix}. \tag{14}$$

We can estimate the parameters in (14) consistently under the null with a GMM type estimator as the one we have described above. Using a block diagonal matrix of instruments Z_i^+, see (13), and under the maintained hypothesis that, using $Z_i^d = diag(x_{i1}, x_{i1}\ x_{i2},\ ...,\ x_{i1}\ x_{i2}\ ...\ x_{iT-1})$, β is consistently estimated under the null *and* the alternative from the first-differences. Then, a Wald test of the null hypothesis that $\alpha_1 = 0$ tests for the lack of correlation between the η_i and the instruments used in Z_i^l. So, if

$$Z_i^l = \begin{bmatrix} x_{i1} & 0 & . & . & 0 & f_i \\ 0 & x_{i3} & . & . & 0 & f_i \\ . & . & . & . & . & f_i \\ . & . & . & . & x_{iT} & f_i \end{bmatrix}$$

a rejection of the null hypothesis in the Wald test can be interpreted as rejecting the null of mean independence[30].

As with the traditional Hausman test we can produce a focused test by only including a set of regressors that take the value of zero in the first-differenced equations and reproducing the level of the variables we are focus on for the equations in levels. If we estimate the system of first-differenced and level equations consistently under the null of mean independence. Then, a Wald test for the null that the coefficients on the additional regressors are jointly zero tests whether the first-differenced estimates of the variables we are focus on (which we maintain are consistently estimated under the null and the alternative) are statistically similar to those obtained using the level equations (consistently estimated only if the instruments in Z_i^l are uncorrelated with the individual effects).

An important feature to notice about Arellano's formulation of the Hausman test is that it not only allows for the fact that the x_{it} can be predetermined or measured with error, but also that the test can be computed using an asymptotic variance-covariance matrix that is robust to heteroskedasticity and autocorrelation of unknown type.

3 Data and results

3.1 Data

The data source used in this work is the British Household Panel Survey (BHPS). The BHPS is a nationally representative survey of some 5,500 households (covering approximately 10,000 individuals) randomly selected. The survey has been conducted from September to December 1991 and annually thereafter [see Taylor (2000)]. Here, we use the eight available waves of the panel. The sample is restricted

[30] Clearly, if $\alpha_2 \neq 0$ and we include f_i in Z_i^l then a Wald test that rejects $\alpha_1 = 0$ will imply that some endogeneity issues are likely to be affecting the estimates of β but not necessarily that $\alpha_1 \neq 0$.

in this study to those individuals who are 16 to 65 years of age, employed in the private sector[31] for paid work, receive a gross hourly wage of more than one pound, work less than a hundred hours a week[32], have complete information on the variables of interest and that do not increase their years of full time education over the sample period. Excluding individuals with less than four consecutive observations, we were left with an unbalanced sample of 1204 individuals with the number of observations varying between four and eight and an average of 5.75 observations per individual.

For reasons of space we will focus on the estimation of wage differentials for time variant characteristics such as employer size, union affiliation and industry status. Finding a suitable set of internal instruments for time invariant characteristics such as years of schooling would merit another paper. Even so, some interesting conclusions and questions for further research on the schooling effect are drawn in this paper. To simplify the exposition that follows, let me call x_{1it} the vector of time varying characteristics that we are interested in (i.e., industry affiliation, union status, and employer size) and x_{2it} those that we are not (i.e., experience[33], experience squared, marital status, type of occupation, part-time status, and location) so that $x_{it} = (x_{1it}, x_{2it})$. Experience is dropped from the set of x_{2it} when we only use equations in first-differences to avoid perfect multicolinearity with the time dummies. The following time invariant characteristics are included when we use conditions in levels: years of full-time schooling, male dummy and white dummy.

Table 1 presents the definition of the variables used, average characteristics, and the number of transitions in and out of a given category that are observed over the entire sample period for some of the variables used to estimate the wage equation. In the next part of this Section we report estimates of size, union, and industry wage differentials. The reference group for each set of dummy variables is, respec-

[31] Private sector is defined in opposition to public sector and it includes: private owned firms, privatized firms, and private non-profit organizations.

[32] Because of the usual omission of labor supply considerations in the estimation of wage equations it is usually advisable to restrict the discussion to full-time employed males. (Of course, unless one is particularly interested in the earning patterns of part-time workers or females.) In our data set we cannot restrict ourselves to full-time employed males as there is not enough time-series variability in that data set to make any sense of the IV techniques we would like to apply. As an attempt to ameliorate any possible misspecification, we introduce in the wage equation a part-time dummy and we interact gender with experience, experience squared and marital status.

[33] As we control for time specific effects by including year dummies we drop experience from the set of time varying covariates when we *only* use moment conditions in first-differences. If we also use moment conditions in levels we do not need to follow this procedure.

tively, service industry, no workplace union, and firm with less than 25 employees. We maintain a common sample period 1993-1998 across all the specifications, the sample contains 1204 individuals and 4515 observations in total. Our estimation procedure was implemented using DPD98 for Gauss [see, Arellano and Bond (1998)] and DPD1.00a for Ox [see, Doornik, Arellano and Bond (1999)].

3.2 Results

Here we present the results of estimating the wage equation, (1), under different sets of assumptions concerning the correlation between right-hand side variables and the error term that we have explained in Section 2.1. Our empirical strategy is to start by estimating the wage equation using the more traditional methods: OLS levels, GLS, within-groups and OLS first-differences. A rule of thumb that emerges from Griliches and Hausman (1986) study of models with errors in variables for panel data is to compare within estimates and OLS differenced estimates. If they turn out to be dissimilar the assumption of strict exogeneity of the x_{it} with respect to v_{is} may be inappropriate. In such cases, we can try to estimate a first-differenced wage equation using suitable lagged and, perhaps, forwarded variables to instrument for the endogenous variables. If we can find a suitable set of instruments for the first-differenced equations, we can try to include moment conditions in levels in order to increase the efficiency of the estimates. Of course, even if we cannot find evidence to reject the strict exogeneity of the x_{it} variables with respect to the time varying idiosyncratic shock we can still benefit from using moment conditions in levels to increase efficiency. If all the x_{it} are uncorrelated with the individual effect, however, there is no need for a fixed-effects transformation. We should take care before reaching this conclusion if we use traditional Hausman test procedures (e.g., Hausman and Taylor (1981)), however, as the correct specification for the test depends on whether the x_{it} are strictly exogenous or not with respect to the v_{is}.

3.2.1 Traditional approaches

We start by presenting in Table 2 the more popular approaches to estimate wage differentials in Labor Economics which, with the time dimension fixed, are only consistent under the assumption of strict exogeneity of the x_{it} with respect to the time varying idiosyncratic shock and/or lack of correlation between the individual effect and $x_{i1}, ..., x_{iT}$ and f_i. The standard errors reported in column (i), (iii) and (iv) are asymptotically robust to time-series or cross-sectional heteroskedasticity and autocorrelation of unknown type.

A necessary condition for a model in levels to provide consistent parameter estimates is that the x_{it} are uncorrelated with the individual effect η_i. In column (i) of Table 2 we present OLS levels estimates. If the necessary condition holds, these are consistent even if x_{it} is only predetermined with respect to v_{is}. The t-statistics and Wald tests suggest that Size, Union and Industry wage differentials are statistically significant at any conventional level. Computing the wage differentials as $\exp(\widehat{\beta})-1$, members of a workplace union earn on average 7.5 percent more than those in jobs where there is no union and individuals in establishments with more than 1000 employees earn on average 34.9 percent more than those working in establishments with less than 25 employees, for example. The m_1 and m_2 tests provide evidence of strong serial correlation, suggesting the presence of individual effects.

If, in addition to strict exogeneity and lack of correlation between the individual effect and the covariates in the wage equation, we assume a typical error component structure with homoskedasticity over time and across individuals. Then, the standard GLS estimator proposed by Balestra and Nerlove (1966) will provide more efficient estimates of the wage differentials than OLS levels. GLS estimates are presented in column (ii). Still, the t-statistics and Wald tests suggest that the wage differentials are statistically significant at any conventional level. In comparison to OLS levels, the GLS standard errors are on average more than forty percent lower suggesting important gains in precision from using GLS. However, the GLS standard errors are not robust to general forms of heteroskedasticity and autocorrelation[34]. Also, notice that the point estimates for size and industry wage differentials are considerably smaller in relation to the OLS levels estimates. For example, in the wage differential for employees in establishments with more than 1000 employees there is a reduction of 17.1 percentage points in relation to OLS levels.

Even under strict exogeneity of the time varying characteristics with respect to v_{is}, the estimates in column (i) and (ii) of Table 2 are inconsistent if the individual effect is correlated with either some of the x_{it} and/or f_i. For a fixed T, under strict

[34] However, these are robust to the type of autocorrelation produced by the presence of the individual effect.

exogeneity, a within group estimator or a first-differenced OLS estimator provides consistent estimates for the effects of time varying characteristics on wages. For size and industry wage differentials, both the OLS levels and GLS estimates are much larger than the within-groups, column (iii), and OLS first-differences estimates, column (iv). In fact, we cannot reject the null that the Industry wage differentials are now jointly statistically insignificant. Comparing the standard errors, there is some loss of precision in comparison to GLS but there is a clear gain in precision either in column (iii) or (iv) in relation to OLS levels. A striking feature from comparing column (iii) and (iv) is the similarity between within-groups and OLS first-differences estimates. This feature also holds for longer lengthened differenced equations estimated by OLS (not reported here). This suggests, after all, that the effect of any biases resulting from the fact that some of the included variables are not strictly exogenous has a minor impact on the estimates of the parameters of interest.

Notice that the union wage differential (especially for members of the workplace union) is fairly similar in the four columns of Table 2. In fact, an heteroskedasticity and autocorrelation consistent Hausman type test (that will only be valid under the assumption of strict exogeneity of the x_{it} with respect to v_{is}) clearly supports the null of no correlation of the union variables with the individual effect. This test is constructed estimating a system of first-differenced and level equations by OLS in which we include a set of regressors that take the value of zero in the first-differenced equations, and reproduce the level of the union variables for the equations in levels. The test statistic is a Wald test for the null that the coefficients on the additional regressors are jointly zero. The value of the χ^2_2 statistic is 0.911 and the p-value is 0.634. However, a similar Hausman test that focuses on all the variables of interest together has a χ^2_{11} value of 87.776 clearly rejecting the null hypothesis that jointly the estimates of the parameters of interests from OLS levels and OLS first-differences are the same.

3.2.2 GMM first-differenced estimates

Although our rule of thumb suggests that the assumption of strict-exogeneity of x_{1it} with respect to v_{is} is perhaps a good idea, we would like to back-up this piece of intuition with some test statistics. In order to do so, we present the results of GMM estimation of first-differenced wage equations under different assumptions concerning the exogeneity of the variables of interest in Tables 3 and 4. This approach is suitable for that enterprise because the first-difference transformation eliminates the individual effect so that we do not need to worry about the correlation between η_i and x_{it}. However, it does not introduce all lagged values of the time varying shocks in the transformed error term so that we can test whether the internal instruments are strictly exogenous or not. The instruments' matrix in Tables 3 and 4 has the general form described in (12). As we would like to focus on testing the validity of the strict-exogeneity assumption for the variables of interest, x_{1it}, we only retain x_{2it-2} and x_{2it-3} in the instruments' set for all columns of Tables 3 and 4. Thus, if there is no evidence of second order serial correlation in the first-differenced residuals the evidence that follows is robust to the lack of strict-exogeneity in x_{2it}. As we go along, we will report the set of instruments involving the variables of interest in each of the columns of Tables 3 and 4. It is worth noticing that theoretically, we could have introduced in the instruments' set further lagged/forwarded variables in many occasions. However, in practice, we are restricted by the fact that the dummy variables do not exhibit enough time-series variation. Moreover, one should expect very distant observations in the past to provide little information on current changes.

In columns (i) to (iv) of Table 3 we present optimal (two-steps) GMM estimates as defined by (11). This estimator allows time varying characteristics to be correlated with the individual effects but assumes that the υ_{it}s are serially uncorrelated. The m_2 statistics reported in columns (i) to (iv) provide a signal that this specification for the υ_{it}s is plausible. Column (i) presents estimates of wage differentials that assume that there is an uncorrelated measurement error structure affecting each of the time varying variables and that the 'true' x_{it} and υ_{is} are correlated for any $s < t$. Under these assumptions only $x_{1it-2}, ..., x_{1i1}$ can be introduced in the instruments' set. In particular, in the estimation of column (i), we only use x_{1it-2} and x_{1it-3}. The Sargan-Hansen test shows that the overidentifying restrictions are not rejected in this case.

In column (ii) we examine whether future instruments may be valid by adding x_{1it+1} and x_{1it+2} to the instruments' set used for column (i). This would be a sound practice if the 'true' x_{1it} were strictly exogenous with respect to υ_{is}. Neither the Sargan-Hansen test rejects the validity of the overidentifying restrictions nor the Sargan-Hansen difference test rejects the validity of the additional instruments (in relation to the set of instruments used in column (i)). This suggests that it is reasonable to assume that the 'true' value of the variables of interest, x_{1it}, are strictly exogenous with respect to the time varying idiosyncratic shock, υ_{is}.

Column (iii) relaxes the measurement error assumption of column (i) and adds x_{1it-1} to the instruments' set used there. Again, neither the Sargan-Hansen test nor the Sargan-Hansen difference test rejects the validity of the additional instruments (in relation to the set of instruments used in column (i)). This suggest that is reasonable to assume that the estimates for the variables of interest, x_{1it}, are not significantly affected by measurement error.

So far there is pretty compelling evidence that the variables of interest are strictly exogenous with respect to the time varying idiosyncratic shocks. Our last test for this hypothesis is produced by adding x_{1it-1}, x_{1it}, x_{1it+1}, and x_{it+2} to the instruments' set of column (i). The Sargan-Hansen test and the Sargan-Hansen difference test support the findings of columns (ii) and (iii).

Monte Carlo evidence reported by Arellano and Bond (1991) and Blundell and Bond (1998) suggest that even with reasonably large sample sizes, inference based on two-step estimators might not be reliable, particularly, when the υ_{it} are heteroskedastic. They recommend that inference (with the exception of Sargan statistics) should be based on first-stage estimators that use robust standard errors. Table 4 reproduces the results of Table 3 using first-stage estimates with robust standard errors.

The m_2 statistics reported in Table 4 are similar to those in Table 3, confirming that it is plausible to assume that the υ_{it}s are serially uncorrelated. In relation to the estimates and their standard errors. In comparison to OLS first-differences, the point estimates in column (i) of Table 4 (the more robust in terms of absence of strict-exogeneity in the x_{1it}) are in general larger in absolute value but their standard errors are also more than three times larger than those obtained using OLS. As we would have expected, there are important gains in precision from increasing the number of instruments we use. For example, expanding the instruments' set of column (i) in Table 4 by adding x_{1it-1}, column (iii), produces an average fall of more than fifty percent in the standard errors.

Not surprisingly, the point estimates and standard errors in column (iv) of Table 4 are very similar to those from within-groups and OLS first-differences in columns (iii) and (iv) of Table 2, respectively. However, the standard errors in column (iv) of Table 3 are half the size of those in the same column of Table 4. This suggest that basing our inference in first-step robust standard errors is probably well justified in this case. Only one out of eight Wald tests computed for Industry and Size differentials can reject, at conventional levels, the null of joint statistical insignificance in Table 4. This is in striking difference with Table 3 where seven out eight Wald test reject the null at either the one or five percent level. Column (iv) of Table 4 is where the larger number of statistically significant wage differentials are detected. The t-test rejects the null that the size dummies and workplace union dummies are individually statistically significant at least at the five percent level. Moreover, the Wald statistic for the Size dummies has a p-value of 0.018.

Before moving on, assuming that in this case the x_{1it} are strictly exogenous with respect to the time varying idiosyncratic shock, we may wonder whether we can trust the evidence we collected from the Sargan-Hansen test. As we know (see, Arellano and Bond (1991) and Ziliack (1997)), this test may tend to under-reject especially in the presence of heteroskedasticity. For this reason we have carried-out a set of simple experiments by including the dependent variable in the instruments' set. We hope that in this way we can explore whether, for our sample size and given the total number of instruments we use, the Sargan-Hansen test can pick-up the presence of any invalid instrument. In addition, this exercise will allow us to learn more about whether the static specification for the wage equation is correct. The results of these experiments are presented in Table 5.

In row (i) of Table 5 we add w_t, w_{t-1}, w_{t-2} and w_{t-3} to the instruments' set used for column (i) in Tables 3 and 4. We know that there are no grounds to introduce w_t and w_{t-1} in the instruments' set. w_{t-2} would be valid instrument only in the absence of serial correlation in the v_{is} and with the model not being incorrectly specified by the exclusion of a lagged dependent variable. In row (ii) we add w_{t-1}, w_{t-2} and w_{t-3} to the instruments' set used for column (i) in Tables 3 and 4. The Sargan-Hansen test clearly rejects the validity of the overidentifying restrictions and the Sargan-Hansen difference test rejects the validity of the additional instruments in both cases. In row (iii) we add w_{t-2} and w_{t-3} to the instruments' set of column (i) in Tables 3 and 4. In this case, neither the Sargan-Hansen test rejects the validity of the overidentifying restrictions nor the Sargan-Hansen difference test rejects the validity of the additional instruments. This suggests that we cannot reject the assumption of lack of serial correlation in the v_{is} and the static specification for the wage equation.

If we repeat this exercise (not reported here) for the instruments' set used in columns (ii) and (iii) of Tables 3 and 4 we find pretty similar results to those reported in rows (i) to (iii) of Table 5. In rows (iv) to (vi) we repeat this exercises using the largest set of instruments' used in the first-differenced equations, those of column (iv) in Tables 3 and 4. In row (iv) the Sargan-Hansen test clearly rejects the validity of the overidentifying restrictions and the Sargan-Hansen difference test rejects the validity of the additional instruments in relation to the set of instruments used in column (i) of Tables 3 and 4. However, in row (v), where we still have the invalid instrument w_{t-1}, the Sargan-Hansen test and the Sargan-Hansen difference test only reject the null marginally, at a ten percent level of statistical significance. This probably suggests that when the set of instruments is relatively large we should take marginal rejections as a serious sign of misspecification for these tests[35].

3.2.3 Using moment conditions in first-differences and in levels

Two closely related issues remain to be investigated. The first is whether any of the x_{1it} can be assumed to be uncorrelated with the individual effects. The second is to what an extent we can increase efficiency by using moment conditions in levels if they were available. In Table 6 we use a GMM estimators which also exploits moment conditions in levels. With the exception of the Sargan-Hansen and Sargan-Hansen difference tests we report first-step robust estimates as we believe that inference based on these estimates is more reliable. Ideally we will use as instruments for the first-differenced equations those in column (iv) of Tables 3 and 4. However, in columns (i) to (iii) we opted for the simplest procedure of using stacked vectors of

[35] We might wonder given the result for the Sargan-Hansen difference test of column (iii) in Table 3 what would have happened if we have only add to the instruments' set in column (iii) x_{1it} rather than x_{1it}, x_{1it+1}, and x_{1it+2} as we did in column (iv). In such case, the p-value for the Sargan-Hansen test is 0.168 and for the Sargan-Hansen difference test is 0.124.

$\triangle x_{it}$ as instruments for the equations in first-differences[36]. We prefer this specification because we are worried about the performance of the Sargan-Hansen statistics with such a number of instruments and there is no much disagreement in the point estimates or standard errors comparing OLS first-differences with column (iv) of Table 4. For the level equations we add to the time varying variables the following time invariant characteristics: years of full-time schooling, male dummy and white dummy. In the instruments' set for the level equations time invariant characteristics are included as strictly exogenous variables and with the same coefficient for all the cross-sections of the first stage regression. It is worth pointing-out that none of the results that follow are affected by excluding schooling from the instruments' set. We will report some schooling estimates in Table 7 although the main focus here is the estimation of the coefficients of x_{1it}. Finally, in concordance to our treatment of the x_{it} variables for the first-differenced equations, we use $\triangle x_{2it}$ as instruments for the level equations in columns (i) to (iii)[37]. As we go along we will report how we included the variables of interest in the instruments' set for columns (i) to (iii) and the complete set of instruments for column (iv).

In column (i) we use as instruments for the level moment conditions first-differences of the variables of interest, $\triangle x_{1it}$. For the extra-moment conditions to be valid, it is only required that the correlation between x_{it} and η_i is assumed constant. The Sargan-Hansen test does not reject the validity of the overidentifying restrictions. Moreover, a one-step robust Hausman-Arellano test for the variables of interest has a p-value of 0.774. This indicates that we cannot reject the null that there are no endogeneity issues affecting the estimates for the variables of interest. In comparison to OLS first-differences the point estimates are similar but there seems to be no gain in precision from using these additional moment conditions.

[36] Obviously, we have only tested the strict exogeneity of the x_{1it} with respect to the time varying idiosyncratic shock. As union status, industry affiliation and employer size are correlated with the x_{2it}, inconsistency in the estimates of the coefficients for the latter variables can carry over to the estimates of the variables of interest. However, for a model in first-differences that uses as instruments x_{it+1}, x_{it} and x_{it-1} the value of the χ^2_{308} statistic for a Sargan-Hansen test is 348.050 which cannot reject the validity of the overidentifying restrictions at a 5 percent level. This imply that we may be in safe ground using a stacked vector of $\triangle x_{it}$ as instruments for the first-differences. Further evidence on the minor impact of this assumption for the estimation of the coefficients of the variables of interest is provided in column (iv) of this Table.

[37] We do not introduce the first differences of the London variable in the instruments' set for the levels equation because it causes perfect multicollinearity. Also, first-differences of experience needs to be excluded from the instruments' set for the levels equation because we are including time dummies.

In column (ii) the instruments' set for the level equations uses the current levels rather than the changes for the variables of interest while we keep the current changes for the time varying variables. If x_{1it} is a valid instrument for the moment conditions in levels, then the x_{1it} are uncorrelated with the individual effect. Both the Sargan-Hansen test and Arellano's version of the Hausman test reject the validity of these instruments, suggesting that at least some of the x_{1it} variables are correlated with the unobservable individual effect. Notice that the estimates for Size and Industry wage differentials increase in relation to OLS first-differences and start looking like the OLS levels estimates. Our tests results suggest that, in general, OLS levels and, in particular, cross-sectional estimates are biased because of the correlation between the individual effects and the right-hand side variables of the wage equation.

Of course, the result in column (ii) does not rule out the fact that some of the variables of interest may in isolation be uncorrelated with the individual effect. Thus, one exercise that we pursue is to mix first-differences and levels of different variables to see whether it is safe to take some of them as uncorrelated with the individual effect. So, we have tried by using in the instruments' set size in levels and union and industry variables in first-differences, union in levels and size and industry variables in first-differences, etc. The only specification that passes all diagnostic tests is presented in column (iii), here we instrument the level equations using the current first-differences for industry and size and the current level only for union. This result confirms our earlier suspicion that the union variable is uncorrelated with the individual effect. The estimate for the union variables might seem a bit higher in relation to OLS first-differences but the standard errors are higher as well. Notice also that, according to the respective Wald statistics, only the size dummies are jointly statistically significant at conventional levels.

A quick look at the standard errors in column (i) to (iii) reveal that there are no efficiency gains in relation to OLS first-differences. Of course, then, we may wonder where the efficiency gains from the level moment conditions have gone. Also we may be interested to know how the system estimates presented so far compare with those that are more robust to the absence of strict-exogeneity in the x_{it} with respect to v_{is}. For this purpose, in column (iv), we estimate a system that uses x_{it-2} and x_{it-3} as instrument for the first-differenced equations (as in column (i) of Table 4). We instrument the level equations using the lagged first-differences for all time varying variables with the exception of union, the level of the union variables lagged once, and stacked values of all the time invariant characteristics[38]. The diagnostic tests do not detect problems with this specification. There is an average gain in precision of around ten percent from the extra moment conditions if we compare with the standard errors of the one-step GMM estimates that only use first-differenced moment conditions (see, column (i) of Table 4). In general, comparing with results in column (iii) of Table 6 the parameter estimates are pretty similar. Notice, however, that if we compare the standard errors in column (iii) with those in column (iv) we find gains in precision of more than fifty percent in column (iii). So, in fact, it is quite reassuring that the instruments' set in column (iii) gives similar point estimates to those obtained with the instruments' set in column (iv) as in the latter we allow *all* the x_{it} to have an uncorrelated measurement error structure and for the 'true' x_{it} to be predetermined with respect to v_{is}.

In Table 7 we present a set of one-step robust GMM estimates for the schooling variable. In column (i) we report the schooling coefficient that we obtained when estimating Table 6. This instruments' set includes stacked values of the schooling variable. As we have seen above, the diagnostic tests provide no evidence of misspecification in the instruments' set used for columns (i), (iii) and (iv). Remarkably, the wage differentials obtained with the instruments' set of columns (i) and (iii) are almost three quarters higher than those from OLS (with almost identical standard errors). In column (ii) of Table 7 we exclude the stacked value of schooling from the instruments' set. Both the point estimates and the standard errors increase slightly in relation to column (i) but these are still very precisely estimated. Not surprisingly, the Sargan-Hansen test does not reject the validity of the overindentifying restrictions in rows (i), (iii) and (iv) of column (ii). In sum, the evidence presented here does not support the view that there is a strong correlation between the schooling variable and the unobserved individual effect. In fact, the OLS levels estimates seem to show considerable downward bias. This is because of inconsistency in other coefficients estimates is carrying over to the schooling coefficient estimates.

[38] See note 37 for the treatment of experience and London variable in the instruments' set for the level equations.

4 Concluding remarks and further research

We have started the paper by pointing out different aspects of the specification of wage equations that should be considered when applying panel data techniques to the estimation of wage differentials. Then, we have presented a family of GMM estimators and specification tests for panel data, with fixed time dimension and large cross-sectional dimension, available under different specifications. Finally, we have applied these techniques to the estimation of a wage equation using the eight available waves of the British Household Panel Survey (BHPS).

We have found that the strict exogeneity assumption of the time varying characteristics with respect to the time varying idiosyncratic shock is not rejected, at least for the variables we focused on this paper (size, union and industry). There is compelling evidence of correlation between the right-hand side variables and the unobserved individual effect in particular for size and industry dummies. This precludes the use of GLS or OLS levels to obtain consistent estimates of any of the coefficients in the wage equation. However, there seems to be no evidence of correlation between the individual effect and the union variables. More surprisingly, none of our diagnostic tests suggest a strong correlation between years of full-time schooling and the individual effect.

The results of our favorite specification (Table 6 column (iii)) reveal that the industry wage differentials are jointly statistically insignificant, once we control for the unobservable individual traits. However, there are statistically significant wage differentials associated with the size of the employer and with the existence of a workplace union. For example, members of a workplace union earn on average 8.9 percent more than those that work in firms where there is no workplace union. The OLS levels estimate for an additional year of schooling is 0.037. The estimate of our preferred specification is 0.064 suggesting that the OLS levels estimates seem to be considerably downward biased because inconsistency in other coefficients estimates is carrying over to the schooling coefficient estimates.

Our findings suggest two important questions for further research. First, we are surprised to find a lack of evidence to support the hypothesis that the variables of interest in this paper are predetermined rather than strictly exogenous. We think that a possible explanation is that we are working with a usual hourly wage measure.

The fact that the reported hourly wage is usual might be smoothing-out some of the feedback and (probably persistence) that one would expect. Unfortunately, it is not possible to construct a point in time hourly wage measure given that only usual weekly hours of work are reported in the BHPS. But it might be worth investigating this issue either with weekly wage data or with other dataset.

Second, most of the literature that studies the schooling wage differential concentrates on controlling for the endogeneity of the schooling variable but omits the problem of potential endogeneity of other right-hand side variables. In particular, this is the case in studies that only use information in levels or in cross-sectional settings (e.g., Angrist and Krueger (1991) and Blundell et. al (2000)). Therefore, our finding that there is considerable bias in the schooling coefficient because of inconsistency in the estimates of other coefficients, rather than because of correlation between schooling and the individual effects, merits further research.

5 References

Amemiya, T. and T.E. MaCurdy, 1986, Instrumental-variables estimation of an error-components model, Econometrica, 54, 869-880.

Angrist, J.D. and A.B. Krueger, 1991, Does compulsory schooling affect schooling and earnings, Quarterly Journal of Economics 106, 979-1014.

Angrist, J.D. and W.K. Newey, 1991, Over-identification tests in earnings functions with fixed effects, Journal of Business & Economic Statistics 9, 317-323.

Arellano, M., 1993, On the testing of correlated effects with panel data, Journal of Econometrics, 59, 87-97.

Arellano, M. and J. Alvarez (1998), The time series and cross-section asymptotics of dynamic panel data models, CEMFI working paper 9808.

Arellano, M. and S.R. Bond, 1991, Some test of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies, 58, 277-297.

Arellano, M. and S.R. Bond, 1998, Dynamic panel data estimation using DPD98- a guide for users.

Arellano, M. and O. Bover, 1995, Another look at the instrumental variable estimation of error-component models, Journal of Econometrics, 68, 5-27.

Balestra, P. and Nerlove, M., 1966, Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas, Econometrica, 34, 585-612.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica, 51, 1635-1659.

Blundell, R.W. and S.R. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics, 87, 29-52.

Blundell, R.W., et al., 2000, The returns to higher education in Britain: Evidence from a British cohort, Economic Journal 110, F82-99.

Bound, J., D.A. Jaeger and R. M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak, Journal of the American Statistical Association 90, 443-447.

Breusch, T., G.E. Mizon and P. Schmidt, 1989, Efficient estimation using panel data, Econometrica 57, 695-700.

Brown, C. and J. Meadoff, 1989, The employer-size wage effect, Journal of Political Economy 97, 1027-59.

Chamberlain, G., 1982, Multivariate regression models for panel data, Journal of Econometrics 18, 5-46.

Chamberlain, G., 1984, Panel data, in: Griliches, Z. and M. Intriligator, Handbook of Econometrics, Volume II, (Elsevier Science Publishers: Amsertdam).

Chowdhury, G. and S.J. Nickell, 1985, Hourly earnings in the United States: Another look at unionization, schooling, sickness, and unemployment using PSID data, Journal of Labor Economics 3, 38-69.

Cornwell, C. and P. Ruppert, 1988, Efficient estimation with panel data: an empirical comparison of instrumental variables estimators, Journal of Applied Econometrics 3, 149-155.

Doornik, J.A., M. Arellano, and S. Bond, 1999, Panel data estimation using DPD for Ox.

Freeman, R.B., 1984, Longitudinal analyses of the effects of trade unions, Journal of Labor Economics 2, 1-26.

Griliches, Z., 1977, Estimating the returns to schooling: Some economtric problems, Econometrica 45, 1-22.

Griliches, Z. and J.A. Hausman, 1986, Errors in variables in panel data, Journal of Econometrics 31, 93-118.

Goux, D. and E. Maurin, 1999, Persistence of Interindustry Wage Differentials: A Reexamination on Matched Worker-Firm Panel Data, Journal of Labor Economics 17, 492-533.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, Econometrica 56, 1371-1395.

Hayashi, F., and C.A. Sims, 1983, Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments, Econometrica 51, 783-98.

Hausman, J.A. and W.E. Taylor, 1981, Panel data and unobservable individual effects, Econometrica 49, 1377-1398.

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

Jakubson, G., 1991, Estimating and testing the union wage effect using panel data, Review of Economic Studies LVII, 971-991.

Keane, M., 1993, Individual heterogeneity and interindustry wage differentials, Journal of Human Resources 28, 134-161.

Keane, M.P. and D.E. Runkle, 1992, On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous, Journal of Business & Economic Statistics, 10, 1-9.

Krueger, A.B., and L.H. Summers, 1988, Efficiency wages and the inter-industry wage structure, Econometrica 56, 259-293.

Mincer, J., 1974, Schooling, experience and earnings (Columbia University Press: New York).

Murphy, K.M. and R.H. Topel, 1987, Unemployment, risk and earnings, in: Lang, K. and J. Leonard, Unemployment & the structure of labor markets (Blackwell: Oxford).

Robinson, C., 1989, The joint determination of union status and union wage effects: Some tests of alternative models, Journal of Political Economy 97, 639-667.

Taylor, M.F. (ed)., 2000, British Household Panel Survey users manual volume A: introduction, technical report and appendices (University of Essex, Colchester).

White, H., 1982, Instrumental variables regression with independent observations, Econometrica, 50, 483-499.

Ziliak, J.P., 1997, Efficient estimation with panel data when instruments are predetermined: An empirical comparison of moment-condition estimators, Journal of Business & Economic Statistics, 15, 419-431.

Table 1: Definitions, average characteristics and number of transitions

Variables		Average characteristics	Number of transitions
Earnings:	Logarithm of usual gross nominal hourly wages	1.829 (0.508)	___
Size:	1 to 24 employees	0.331 (0.471)	616
	25 to 199 employees	0.359 (0.480)	900
	200 to 999 employees	0.226 (0.418)	663
	1000 or more employees	0.083 (0.276)	255
Union:	Member of work-place union	0.243 (0.429)	227
	Non-member of work-place union	0.137 (0.344)	361
	No work-place union	0.620 (0.485)	372
Industry:	Mining	0.007 (0.082)	26
	Construction	0.026 (0.159)	62
	Manufacturing	0.346 (0.476)	364
	Transport, Utilities & Communication	0.097 (0.296)	155
	Wholesale & retail trade	0.192 (0.394)	337
	Finance, Insurance & Real Estate	0.127 (0.333)	89
	Services	0.205 (0.404)	323
Marital status:	Married	0.682 (0.466)	189
Part time:	Work less than 35 hours a week	0.168 (0.374)	226
Schooling:	Years of full time education	12.422 (2.647)	0
Experience:	Age minus 5 minus schooling	21.551 (11.709)	___
Location:	Southeast	0.196 (0.397)	33
	London	0.097 (0.296)	19
Type of Occupation:	Professional	0.072 (0.259)	254
	Managerial or technical	0.319 (0.466)	705
	Skilled	0.481 (0.500)	818
	Unskilled	0.128 (0.334)	328
Gender:	Male	0.566 (0.496)	0
Ethnic background:	White	0.978 (0.147)	0
Total Number of Individuals		1204	
Total Number of Observations		6923	

Source: British Household Panel Survey (BHPS).
Notes:
1) Standard deviations are reported in parentheses.
2) Transitions are in and out of a given category.

Table 2: OLS levels, GLS, within-groups and OLS first-differences estimates

Dependent variable: Logarithm of usual gross nominal hourly wages	(i)	(ii)	(iii)	(iv)
25 to 199 employees	0.148**	0.063**	0.025*	0.020
	(0.022)	(0.011)	(0.013)	(0.013)
200 to 999 employees	0.171**	0.081**	0.034*	0.032*
	(0.024)	(0.014)	(0.015)	(0.015)
1000 or more employees	0.299**	0.164**	0.082**	0.094**
	(0.034)	(0.020)	(0.024)	(0.029)
Member of work-place union	0.072**	0.072**	0.054**	0.062**
	(0.021)	(0.015)	(0.017)	(0.019)
Non-member of work-place union	0.030	0.039**	0.014	0.025
	(0.022)	(0.013)	(0.016)	(0.016)
Mining	0.423**	0.112*	0.060	0.093*
	(0.107)	(0.052)	(0.039)	(0.037)
Construction	0.033	0.003	0.006	0.039
	(0.055)	(0.037)	(0.047)	(0.041)
Manufacturing	0.049	0.001	0.007	0.005
	(0.028)	(0.017)	(0.025)	(0.022)
Transport, Utilities & Communication	0.055	0.003	-0.006	0.004
	(0.039)	(0.022)	(0.029)	(0.028)
Wholesale & Retail Trade	-0.068*	-0.079**	-0.041	-0.026
	(0.032)	(0.018)	(0.027)	(0.024)
Finance, Insurance & Real Estate	0.255**	0.133**	-0.003	0.002
	(0.040)	(0.023)	(0.038)	(0.046)
Years of full time schooling	0.037**	0.056**	—	—
	(0.005)	(0.004)		
m_1	12.451	8.067	-27.119	-6.536
(*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)
m_2	10.475	2.763	2.684	1.647
(*p*-value)	(0.000)	(0.000)	(0.007)	(0.100)
Wald Size	89.694	77.475	12.341	10.431
(*p*-value)	(0.000)	(0.000)	(0.006)	(0.015)
Wald Industry	81.184	83.824	10.302	11.783
(*p*-value)	(0.000)	(0.000)	(0.112)	(0.067)
Method of Estimation	OLS Levels	GLS	Within-Groups	OLS First-Differences

Source: British Household Panel Survey (BHPS).

Notes:

1) Sample period is 1993-1998. There are 1204 individuals and 4515 observations.

2) Asymptotic standard errors are reported in parentheses.

3) In columns (i), (iii) and (iv) standard errors are robust to general time-series and cross-sectional heteroskedasticity .

4) ** statistically significant at a 1 percent level. * statistically significant at a 5 percent level.

5) Other regressors included in the wage equations. Columns (i) and (ii): Experience interacted with Gender, Experience Squared interacted with Gender, Male dummy, White dummy, Marital status dummy interacted with Gender, Time dummies, London dummy, South east dummy, Part-time dummy, Professional occupation dummy, Managerial and technical occupation dummy, and Unskilled dummy. Columns (iii) and (iv): as in columns (i) and (ii) but excluding Experience interacted with Gender, Schooling, Male dummy, and White dummy.

6) The omitted dummies of interest are less than 25 employees, No workplace union and Services industry.

7) m_2 and m_1 are tests of first and second order serial correlation in first differenced disturbances, asymptotically distributed as N(0,1) under the null of no serial correlation.

8) Wald is a test of joint statistical significance of the corresponding dummies, asymptotically distributed Chi-square under the null with 3 degrees of freedom for Size and 6 for Industry.

9) In column (i), (iii) and (iv) estimation was implemented using DPD98 for Gauss (see, Arellano and Bond (1998)) and in column (ii) using DPD1.00a for Ox (Doornik, Arellano and Bond (1999)).

Table 3: Optimal (two-steps) first-differenced GMM estimates

Dependent variable: Logarithm of usual gross nominal hourly wages	(i)	(ii)	(iii)	(iv)
25 to 199 employees	0.038	0.058**	-0.014	0.030**
	(0.030)	(0.019)	(0.012)	(0.006)
200 to 999 employees	0.043	0.050*	0.008	0.042**
	(0.038)	(0.023)	(0.017)	(0.007)
1000 or more employees	0.024	0.049**	0.040	0.065**
	(0.056)	(0.034)	(0.022)	(0.010)
Member of work-place union	0.130**	0.081**	0.109**	0.049**
	(0.038)	(0.024)	(0.020)	(0.009)
Non-member of work-place union	0.077**	0.060	0.049**	0.033**
	(0.027)	(0.021)	(0.014)	(0.008)
Mining	0.119*	0.060	0.038	0.063**
	(0.059)	(0.056)	(0.030)	(0.022)
Construction	-0.158	-0.063	0.023	0.037**
	(0.099)	(0.051)	(0.024)	(0.010)
Manufacturing	-0.115*	-0.019	-0.035	0.023*
	(0.058)	(0.038)	(0.020)	(0.010)
Transport, Utilities & Communication	-0.136	-0.087	-0.036	0.001
	(0.081)	(0.048)	(0.025)	(0.013)
Wholesale & Retail Trade	0.018	0.032	-0.018	-0.021
	(0.072)	(0.042)	(0.022)	(0.012)
Finance, Insurance & Real Estate	0.038	0.112*	0.077**	0.000
	(0.071)	(0.051)	(0.039)	(0.013)
m_1	-6.742	-6.634	-6.597	-6.552
(*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)
m_2	0.949	1.372	1.334	1.492
(*p*-value)	(0.342)	(0.170)	(0.182)	(0.136)
Wald Size	1.938	9.593	8.536	51.951
(*p*-value)	(0.371)	(0.022)	(0.036)	(0.000)
Wald Industry	21.312	19.181	17.415	49.800
(*p*-value)	(0.002)	(0.004)	(0.008)	(0.000)
Sargan-Hansen [d.f.]	216.144 [210]	308.343 [309]	293.381 [276]	441.647 [441]
(*p*-value)	(0.371)	(0.500)	(0.226)	(0.482)
Sargan-Hansen Difference [d.f.]	___	92.196 [99]	77.273 [66]	225.503 [231]
(*p*-value)		(0.673)	(0.162)	(0.590)
Instruments	X_{1i}: *t-2, t-3*	X_{1i}: *t+2, t+3, t-2, t-3*	X_{1i}: *t-1, t-2, t-3*	X_{1i}:*t+2, t+1, t, t-1, t-2, t-3*
	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*

Source: British Household Panel Survey (BHPS).

Notes:

1) Sample period is 1993-1998. There are 1204 individuals and 4515 observations.

2) Asymptotic standard errors are reported in parentheses. Standard errors are robust to general time-series and cross-sectional heteroskedasticity.

3) ** statistically significant at a 1 percent level. * statistically significant at a 5 percent level.

4) Other regressors included in the wage equations: Experience Squared interacted with Gender, Marital status dummy interacted with Gender, London dummy, South east dummy, Part-time dummy, Professional occupation dummy, Managerial and technical occupation dummy, and Unskilled dummy. Time dummies are included as regressors and instruments in all equations.

5) The omitted dummies of interest are less than 25 employees, No workplace union and Retail and wholesale trade.

6) m_2 and m_1 are tests of first and second order serial correlation in first differenced disturbances, asymptotically distributed as N(0,1) under the null of no serial correlation.

7) Wald is a test of joint statistical significance of the corresponding dummies, asymptotically distributed Chi-square under the null with 3 degrees of freedom for Size and 6 for Industry.

8) Sargan-Hansen is a test of overidentifying restrictions, asymptotically distributed Chi-square under the null with degrees of freedom in brackets.

9) Sargan-Hansen difference is a nested test for the validity of the additional instruments, asymptotically distributed Chi-square under the null with degrees of freedom in brackets. The nested instruments' set is always the one we use for column (i).

10) Estimation was implemented using DPD98 for Gauss (see, Arellano and Bond (1998)).

Table 4: One-step robust first-differenced GMM estimates

Dependent variable: Logaritm of usual gross nominal hourly wages	(i)	(ii)	(iii)	(iv)
25 to 199 employees	0.010	0.043	-0.011	0.024*
	(0.041)	(0.029)	(0.020)	(0.012)
200 to 999 employees	-0.026	0.001	-0.011	0.032*
	(0.056)	(0.040)	(0.028)	(0.014)
1000 or more employees	-0.047	-0.007	0.038	0.075**
	(0.092)	(0.064)	(0.039)	(0.025)
Member of work-place union	0.096	0.068	0.087*	0.071**
	(0.059)	(0.039)	(0.039)	(0.018)
Non-member of work-place union	0.060	0.025	0.036	0.030**
	(0.039)	(0.032)	(0.023)	(0.014)
Mining	0.209	0.172	0.123*	0.087
	(0.172)	(0.095)	(0.058)	(0.036)
Construction	-0.148	-0.039	0.014	0.025
	(0.172)	(0.119)	(0.052)	(0.042)
Manufacturing	-0.087	0.008	-0.014	0.014
	(0.088)	(0.063)	(0.033)	(0.023)
Transport, Utilities & Communication	-0.088	-0.008	-0.001	0.000
	(0.121)	(0.082)	(0.045)	(0.028)
Wholesale & Retail Trade	0.090	0.063	0.016	-0.024
	(0.105)	(0.078)	(0.036)	(0.024)
Finance, Insurance & Real Estate	0.069	0.132	0.076	-0.007
	(0.102)	(0.084)	(0.079)	(0.039)
m_1	-6.662	-6.578	-6.541	-6.562
(*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)
m_2	0.909	1.320	1.304	1.632
(*p*-value)	(0.363)	(0.187)	(0.192)	(0.103)
Wald Size	0.858	3.168	2.579	10.091
(*p*-value)	(0.836)	(0.366)	(0.461)	(0.018)
Wald Industry	7.733	7.874	7.585	11.145
(*p*-value)	(0.258)	(0.247)	(0.270)	(0.084)
Instruments	X_{1i}: *t-2, t-3*	X_{1i}: *t+2, t+3, t-2, t-3*	X_{1i}: *t-1, t-2, t-3*	X_{1i}:*t+2, t+1, t, t-1, t-2, t-3*
	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*	X_{2i}: *t-2, t-3*

Source: British Household Panel Survey (BHPS).
Notes: As in Table 3.

Table 5: Test-specification for first-differenced GMM estimates: Including the dependent variable in the instruments' set

Additional Instruments:	Sargan-Hansen [d.f.] (*p*-value)	Sargan-Hansen Difference [d.f.] (*p*-value)	Base Instruments' set
(i) w_t , w_{t-1}, w_{t-2}, w_{t-3}	311.786 [233]	95.642 [23]	
	(0.000)	(0.000)	
(ii) w_{t-1}, w_{t-2}, w_{t-3}	284.872 [227]	68.728 [17]	X_{1i}: *t-2, t-3*
	(0.005)	(0.000)	X_{2i}: *t-2, t-3*
(iii) w_{t-2}, w_{t-3}	234.928 [221]	18.784 [11]	
	(0.248)	(0.065)	
(iv) w_t , w_{t-1}, w_{t-2}, w_{t-3}	563.339 [464]	347.195 [254]	
	(0.001)	(0.000)	
(v) w_{t-1}, w_{t-2}, w_{t-3}	498.800 [458]	282.656 [248]	X_{1i}:*t+2, t+1, t, t-1, t-2, t-3*
	(0.091)	(0.064)	X_{2i}: *t-2, t-3*
(vi) w_{t-2}, w_{t-3}	463.737 [452]	247.593 [242]	
	(0.341)	(0.389)	

Source: British Household Panel Survey (BHPS).

Notes:

1) Sample period is 1993-1998. There are 1204 individuals and 4515 observations.

2) Regressors included in the wage equations: 3 Size dummies, 2 work-place union dummies, 6 Industry dummies Experience Squared interacted with Gender, Marital status dummy interacted with Gender, London dummy, South east dummy, Part-time dummy, Professional occupation dummy, Managerial and technical occupation dummy, and Unskilled dummy. Time dummies are included as regressors and instruments in all equations.

3) Sargan-Hansen is a test of overidentifying restrictions, asymptotically distributed Chi-square under the null with degrees of freedom in brackets.

4) Sargan-Hansen difference is a nested test for the validity of the additional instruments, asymptotically distributed Chi-square under the null with degrees of freedom in brackets. The nested instruments' set is always the one we use for column (i) of Tables 3 and 4.

5) Estimation was implemented using DPD98 for Gauss (see, Arellano and Bond (1998)).

Table 6: One-step robust first-differenced/levels GMM estimates

Dependent variable: Logarithm of usual gross nominal hourly wages	(i)	(ii)	(iii)	(iv)
25 to 199 employees	0.025	0.132**	0.025	0.028
	(0.015)	(0.019)	(0.015)	(0.044)
200 to 999 employees	0.038*	0.182**	0.040*	0.106
	(0.019)	(0.022)	(0.018)	(0.058)
1000 or more employees	0.100**	0.309**	0.106**	0.126
	(0.032)	(0.032)	(0.032)	(0.094)
Member of work-place union	0.073**	0.020	0.085**	0.117**
	(0.026)	(0.022)	(0.024)	(0.032)
Non-member of work-place union	0.038*	0.006	0.061**	0.073*
	(0.019)	(0.022)	(0.023)	(0.033)
Mining	0.099	0.283**	0.104	0.304
	(0.062)	(0.092)	(0.063)	(0.193)
Construction	-0.020	0.018	-0.018	-0.055
	(0.046)	(0.047)	(0.045)	(0.138)
Manufacturing	-0.006	0.001	-0.002	-0.014
	(0.027)	(0.028)	(0.027)	(0.074)
Transport, Utilities & Communication	0.017	0.022	0.012	-0.033
	(0.034)	(0.039)	(0.033)	(0.089)
Wholesale & Retail Trade	-0.033	-0.097**	-0.037	-0.019
	(0.029)	(0.032)	(0.029)	(0.083)
Finance, Insurance & Real Estate	0.018	0.263**	0.030	0.035
	(0.047)	(0.037)	(0.047)	(0.109)
m_1	-6.620	-7.032	-6.624	-7.269
(*p*-value)	(0.000)	(0.000)	(0.000)	(0.000)
m_2	1.550	1.622	1.566	1.286
(*p*-value)	(0.121)	(0.105)	(0.117)	(0.199)
Wald Size	10.051	108.895	11.612	3.909
(*p*-value)	(0.018)	(0.000)	(0.009)	(0.272)
Wald Industry	6.953	98.054	8.001	3.187
(*p*-value)	(0.325)	(0.000)	(0.237)	(0.785)
Sargan-Hansen [d.f.]	151.248 [138]	210.733 [149]	162.910 [140]	338.520 [330]
(*p*-value)	(0.208)	(0.001)	(0.090)	(0.361)
Sargan-Hansen Difference [d.f.]	___	___	___	122.376 [120]
(*p*-value)				(0.423)
Hausman-Arellano	7.308	93.124	9.571	16.881
(*p*-value)	(0.774)	(0.000)	(0.569)	(0.111)
Instruments	*Dif:stacked* ΔX_{it}	*Dif:stacked* ΔX_{it}	*Dif:stacked* ΔX_{it}	*Dif:* X_{1i}*: t-2, t-3* X_{2i}*: t-2, t-3*
	Lev: ΔX_{1it}, ΔX_{2it}	*Lev:* X_{1it}, ΔX_{2it}	*Lev:* $Union_{1it}$, *all other* ΔX_{it}	*Lev:* $Union_{1it-1}$, *all other* ΔX_{it-1}

Source: British Household Panel Survey (BHPS).

Notes:

1) Sample period for first-differenced equations is 1993-1998. There are 1204 individuals and 4515 observations.

2), 3) See the corresponding notes in Table 3.

4) Other regressors included in the first-differenced equations: Experience Squared interacted with Gender, Marital status dummy interacted with Gender, London dummy, South east dummy, Part-time dummy, Professional occupation dummy, Managerial and technical occupation dummy, and Unskilled dummy. Other regressors included in the level equations: the same used for the first differences plus Experience interacted with Gender, Schooling, Male dummy, White dummy. Time dummies are included as regressors and instruments in all equations. In the instruments' set for the level equations neither first-differences in experience or the London dummy are included. Male dummy , White dummy and Schooling are included stacked in the instruments' set for the level equations.

5), 6), 7), 8) See the corresponding notes in Table 3.

9) Sargan-Hansen difference is a nested test for the validity of the additional instruments, asymptotically distributed Chi-square under the null with degrees of freedom in brackets. The nested instruments' set is always the one we use for column (i) of Tables 3 and 4.

10) Hausman-Arellano is Arellano's version of the Hausman test which is computed from on an auxiliary regression that includes another set of regressors that take the value of zero in the equation in first-differences, and reproduce the levels of the variables of interest for the equations in levels. The test statistic is a Wald test of the hypothesis that the coefficients on these additional regressors are jointly zero, asymptotically distributed Chi-square under the null with 11 degrees of freedom.

11) Estimation was implemented using DPD98 for Gauss (see, Arellano and Bond (1998)).

Table 7: Some one-step robust first-differenced/levels GMM estimates for the schooling coefficient

Status of Schooling in the instruments' set for the level equations:	Included (i)	Excluded (ii)	
Base Instruments' set:	Schooling Coefficient (s.e.)	Schooling Coefficient (s.e.)	Sargan-Hansen [d.f.] (*p*-value)
Dif: stacked ΔX_{it}	0.063**	0.076**	151.520 [137]
Lev: ΔX_{1it}, ΔX_{2it}	(0.006)	(0.016)	(0.187)
Dif: stacked ΔX_{it}	0.059**	0.093**	199.194 [148]
Lev: X_{1it}, ΔX_{2it}	(0.006)	(0.014)	(0.003)
Dif: stacked ΔX_{it}	0.064**	0.071**	163.377 [139]
Lev: $Union_{1it}$, all other ΔX_{it}	(0.006)	(0.016)	(0.077)
Dif: X_{1i}: t-2, t-3, X_{2i}: t-2, t-3	0.052**	0.044**	338.357 [329]
Lev: $Union_{1it-1}$, all other ΔX_{it-1}	(0.007)	(0.017)	(0.349)

Source: British Household Panel Survey (BHPS).
Notes:
1) Sample period is 1993-1998. There are 1204 individuals and 4515 observations.
2) Asymptotic standard errors are reported in parentheses. Standard errors are robust to general time-series and cross-sectional heteroskedasticity.
3) ** statistically significant at a 1 percent level. * statistically significant at a 5 percent level.
4) For column (ii), other regressors included in the first-differenced equations are: 6 industry dummies, 2 work-place union dummies, 3 size dummies, Experience Squared interacted with Gender, Marital status dummy interacted with Gender, London dummy, South east dummy, Part-time dummy, Professional occupation dummy, Managerial and technical occupation dummy, and Unskilled dummy. Other regressors included in the level equations: the same used for the first differences plus Experience interacted with Gender, Schooling, Male dummy, White dummy. Time dummies are included as regressors and instruments in all equations. In the instruments' set for the level equations neither first-differences in experience or the London dummy are included. Male dummy and White dummy are included stacked in the instruments' set for the level equations. Column (i) reports the estimates for the schooling variables obtained in column (i) to (iv) of Table 6. Therefore, the only difference with Column (ii) of this Table is the inclusion of the schooling variable stacked in the instruments' set for the level equations.
5) Sargan-Hansen is a test of overidentifying restrictions, asymptotically distributed Chi-square under the null with degrees of freedom in brackets. See Table 6 for the Sargan-Hansen tests for column (i).
6) Estimation was implemented using DPD98 for Gauss (see, Arellano and Bond (1998)).