The Wage Effect of Working in the Public Sector When Education and Sector Choices Are Endogenous: An Empirical Investigation for Italy

Paolo Ghinetti[♥]
(address for correspondence: paolo.ghinetti@unicatt.it)

Abstract: This paper estimates the public-private wage differentials for Italy using a pooled cross-section from 1995, 1998 and 2000 household survey and by assuming endogeneity of education and sector choices. Methods to estimate the pay premium under constant or individual-specific effects are discussed. Results show that returns to education are higher than OLS predictions due to negative selection in education. In addition, neglecting the importance of its endogenous nature produces an upward bias in the wage premium. For what concerns sector choices, public employees work into the sector in which their (unobserved) productivity is lower. Because of negative self-selection, the wage premium is positive (8.5%) but lower than for a random individual (14.5%).Accordingly, the choice to work in the public sector seems to be explained by more favourable working conditions and not by a comparative wage advantage.

JEL codes: J31, J45, C3.
Keywords: public wage premium, Microeconometrics, Italy.

[♥] I am indebted to L. Cappellari, C. Lucifora and A. Kugler for very helpful discussions and to J. Wooldridge for valuable suggestions. Useful comments were also received from participants at Reading groups in Labour microeconometrics - Università Cattolica and 7th IZA European Summer School in Labour Economics, and in particular from S. Comi, F. Origo, E. Melero, A. Arellano and M. Sylos Labini. Data were kindly provided by the Bank of Italy and are freely downloadable from the website of the Institution. Usual disclaimers apply.

1. Introduction

In many industrialised countries the state accounts for a significant share of total employment and, both by producing good and services and by regulating the activity of the private sector, deeply influences the functioning of the entire economy. In this context, Italy is not an exception, and public intervention in the economy is substantial, and occurs at the central (government) and at the local level, as well as by means of public authorities. By their nature, goods and services offered by the public sector are essential and (often) produced from a monopolistic position, and rules governing employment conditions, human resource management and pay determination are intrinsically different across public and private sectors. For example, in Italy large differences exist in recruitment, retention and incentive policies, as well as in careers and wage profiles. Moreover, the availability of most occupation is not the same across the two sectors. In the public sector, once hired through a public examination, a civil servant enjoys a lifetime working contract and seniority plays a key role in wage progression. In addition, incentives relating wages to productivity are often missing. In the private sector, although the power of "insiders" is still high and wages are - for the most part - collectively negotiated, the degrees of flexibility in wage determination are higher – bargaining in Italy takes place also at the industry level - and the criteria to hire and promote workers are less strict than in the public sector. In addition, as a consequence of higher union power and because often the State has some interest to be perceived as a "good employer" by offering (relatively) high wages to low skilled workers and (relatively) low wages to the high skilled, the wage structure in the public sector tend to be flat compared to the private sector.

Despite the evidence that wages and working conditions greatly differ between the public and private sector, only recently economists have become interested in analysing the wage effect of working in the public sector. Starting from Smith (1977), the literature typically focussed on the estimation of (conditional) structural differences between public and private wages. For the most part, US studies report that the public sector has a less elastic labour demand curve and that rents are being earned by public sector employees with respect to private sector workers with the same (observable) characteristics.

However, as a consequence of differences in job attributes, working condition and hiring requirement, the choice of the sector might not be random. In particular, some employees may exhibit a preference for working in the public sector, self-selecting themselves according to unobservable characteristics. Since in this context the OLS estimator is biased and inconsistent, a number of studies for different countries estimated wage regressions using maximum likelihood and two-step Heckman methods to compute public pay premium free of selection bias. Examples are Hartog and Oosterbeek (1993) for Netherlands, Belman and Heywood (1989) for the US, Dustmann and Van Soest (1998) for Germany, Disney and Gosling (1998) for UK and Adamchick and Bedi (2000) for Poland. As a consequence of differences in wage structures, institutional settings and workforce selection mechanisms across countries, results from these studies show a great deal of variation in the estimated premium.

For what concerns Italy, OLS estimates of the public-private wage differential (Cannari et al, 1989: Brunello and Rizzi, 1993; Brunello and Dustmann, 1997; Lucifora, 1999; Comi and Ghinetti, 2002) varies in the range of 9-12% depending on the period considered, the sample used, the specification adopted and the definition of public sector employed. Estimates obtained controlling for endogenous sector choices vary considerably more than those obtained from OLS: using the single equation model with an endogenous dummy for sector affiliation Cannari et al. (1989) and Brunello and Rizzi (1993) find that the (conditional) wage differential is not significantly different from zero (and negative for males (-23%); positive (7%) for women); using the more flexible specification with two equations and endogenous switching Brunello and Dustmann (1997) report that, at least for males, the premium is positive (21%) and it can be largely explained by observable workers' attributes, while Bardasi (1996), who uses a more sophisticated selection procedure – workers can choose to work in the public sector, in the private sector or to be self-employed -, finds that the observed differential is substantial for women (35%) and smaller for men (8.8 %), and that the larger contribution (40% for male, 50% for women) comes from different returns paid to similar characteristics and that the effects of different (observed) characteristics is not significant[1].

Also the magnitude of selection effects varies considerably across studies: according to Cannari et al (1989) and Brunello and Rizzi (1993) they are weak, whereas Brunello and Dustmann (1997) find no evidence of such effects; Bardasi (1996) reports that significant and negative endogenous selection exists in the public-private occupational choice.

On the one hand, differences in estimates across studies may reflect the high volatility of the wages in both sectors over the period considered (end of 80s, beginning of 90s), which is disturbing when using cross-sections from different years. On the other hand, these

[1] Cappellari (2002) takes an alternative route to the approach based on static differences in earnings between the two sectors and investigates the dynamic of earnings. He finds that life cycle considerations matter in the formation of the differential; in the private sector careers are less stable and the growth rate of wages is more volatile that in the public sector, where wages are more homogeneous over the life.

differences may depend on the sensitivity of results to model assumptions and identification strategies.

For what concerns model assumptions, a common feature of these studies is to consider the mechanism of self-selection into the public sector as the unique source of endogenous selection. Education, which enters as a determinant of both the wage and the sector choice, is always treated as exogenous. Considering education as exogenously determined is problematic: as pointed out by the vast literature on returns to schooling and confirmed by recent studies for Italy – see, for example, Brunello and Miniaci (1999) and Colussi (1997) -, educational attainment is correlated with unobservable wage determinants (and in particular ability) and, therefore, is likely to be endogenous in the wage equation. This (potential) problem may bias not only the return to education, but also the estimation of other parameters, including the public wage premium, especially when (unobservable) preferences for the public sector are correlated with (unobservable) variables driving decisions about the optimal level of education[2]. In Italy, where the public sector recruits through concourses in which the level of education plays a key role, this problem may be relevant and, therefore, existing studies may report biased and inconsistent estimates of the wage premium.

For what concerns identification, sector choices of individuals with a given set of (observed and unobserved) personal characteristics and preferences can be adequately modelled only if the selection equation is correctly specified. This point is crucial, since in many cases identification heavy relies instruments whose validity may be questionable since it is not clear whether these variables are excludable from the wage equation.

In addition, number of studies assume that unobserved wage determinants among participants are equal in both sectors, and the public premium constant and not individual-specific. But, since working conditions and job attributes differ substantially between public and private occupations, unobservable determinants of wages may differ across sectors, and public employees may be self-selected on the basis of these differences.

The aim of this paper is to estimate the public-private wage differential in Italy using a pooled cross section of observations from 1995, 1998 and 2000 Bank of Italy's Survey of Households Income and Wealth.

With respect to the existing literature, several novelties are offered. First of all, I present an extension of the Heckman (1978) specification of a (wage) regression with an endogenous dummy variable by allowing that not only the sector of employment, but also the education attainment are endogenous to the wage and potentially correlated each other.

Second, the validity of identification strategies for the two endogenous variables is carefully investigated and tested. Finally, as an alternative to the model with constant unobserved heterogeneity, I also (derive and) estimate an extended version of the model where the effect of working in the public sector is individual-specific. In other words, contrary to the original model, where the wage differential is measured by the coefficient of the sector dummy and therefore is constant in the population, in the extended model a public worker and a randomly chosen individual receive a different wage premium because the former obtain random wage gains (or losses) when working in the sector in which he is actually employed.

The paper is organised as follows. In section 2 the main features of the data are described. Section 3 introduces the econometric framework and discuss, under different assumptions, estimation techniques and the identification strategies. Main results are offered in section 4. Conclusions follow in section 5.

2. Data and Variables

Data used in this paper are drawn from the 1995,1998 and 2000 waves of the Survey of Household Income and Wealth (SHIW). Each wave is based on a random sample of around 20,000 individuals. Although the sampling unit is the household, detailed information is available also at individual level, like education, gender, age, work experience, region of residence and of origin, occupation, (net and gross) yearly earnings, average weekly hours of work and number of months of employment per year. The Survey provides also detailed family background information, like parents' education, occupation, sector of employment. The sample used in the empirical investigation is drawn by the population of non-agricultural workers who are in dependent employment and aged from 15 to 65[3]. Using this information, I construct a pooled cross-section for the years of interest. The pooling procedure is used to improve the asymptotic properties of the estimates and is not rejected by the data. Earlier waves has not been used because bargaining procedures until 1993 were significantly different from those adopted in the subsequent period. For the subsequent analysis, there are two potential shortcomings. First, the definition of the public sector refers to the Italian "*Pubblica Amministrazione*", which excludes firms financed by the state which operate in the market. To mitigate this problem, a worker has been classified as a public employees using information from the variable "firm size" – which in the sector presents a specific

[2] See Dustmann and Van Soest (1998) for evidence on the fraction of education and sector choices simultaneously determined.

[3] Due to data limitations on family background information, the sample is further restricted to households, spouses and sons of the household.

classification[4]. Second, no information is available on the number of weeks worked on average in a month. According to all previous studies (see Bardasi, 1996, for a detailed discussion over this issue), hourly earnings are computed (at 1995 prices) assuming that an individual worked 52/12 weeks per month, and are inclusive of extra-time compensations and fringe benefits, and net of taxes and social security contributions. For the three years of interest, summary statistics are given in Table 1[5].

< TABLE 1 AROUND HERE>

Public sector employment accounts for more that 30% of the sample. On average, public sector employees are older (43 years) and hold higher levels of education (12 years) than private sector employees (37 and 10 years respectively). The fact that specific levels of education attainment are usually required to obtain public sector positions may explain this pattern. Hourly (unconditional) average wages are higher in the public sector than in the private sector and they show also a lower dispersion. The raw differential is about 29% and is statistically significant[6].

3. The Econometric Methodology

In order to investigate whether individuals receive an equal remuneration in the two sectors or not, the simple comparison of wages between the public and the private sector does not provide enough information. The computation of the wage differential requires knowledge about the wage that an individual working in the public sector would receive in the private sector, maybe controlling for other determinants besides the sector. As the same person cannot be in two different labour market states at the same time and we observe the wage only for the sector in which a worker is actually employed, the counterfactual situation is not observable and can only be estimated using information on private sector's workers.

Let "working in the public sector" to be the treatment (D) received by an individual, and his or her wage the outcome of that treatment. Private sector employees are the "control group". The parameter of interest is the "Average Treatment effect on the Treated" (ATT), which is the mean difference between the wage actually earned by public sector employee and his (potential and not observed) wage in the private sector (counterfactual situation). If assignment to sectors is not exogenous, the wage received by (comparable) workers in the private sector is not necessarily a good estimator of the wage earned by public employees had

[4] Results from estimates obtained without the correction do not significantly differ from those reported in section 4 and are available upon request.
[5] The data were cleaned by excluding outliers and missing values for relevant variables. There are no reasons to belive that excluded individuals have any systematic relationship with the these variables.
[6] The p-value of a t-test for zero mean difference is 0.000.

they worked in the other sector: persons who work in the public sector are different from persons who do not, in the sense that mean outcomes of participants in the non participation state would be different of those of non participants. Two main reasons are responsible for that. The first one is the traditional endogeneity problem: this happens when the wage impact of the public sector is constant across individuals but public sector workers have *on average* a lower unobservable productivity than otherwise similar individuals working in the private sector. If productivity levels were correlated with the decision to work in the public sector and, at the same time, affected the wage received in both sectors, the public sector workers would earn less than otherwise similar workers, had they worked in the private sector. Failure to control for this difference would lead the lower wage of those with lower ability working in the public sector to be incorrectly attributed to their sector affiliation.

The second reason is self-selection: let assume that the wage impact of working in the public sector is different across individuals and function of unobservable variables that, in turn, also affect the probability to be either a public or a private employee. Then, selection into the two sectors sector may not be random. If estimates come from selected samples, one might expect that the true average treatment effect for treated individuals is different than the average treatment effect estimated using information on self-selected private employees.

In both cases (endogeneity and self-selection) OLS estimates are biased and inconsistent. To solve these problems, Heckman (1979) and Heckman and Robb (1985) developed a two-step procedure which, under specific distributional assumptions, consistently estimate endogenous treatment effects by modelling the stochastic dependence between the unobserved determinants of the outcome and the endogenous treatment. This dependence usually takes the form of controls functions (correction terms) known up to some estimable parameters.

However, if additional covariates besides sector assignment are correlated with unobservable determinants of wages - and, also, with preferences for the public sector – standard two-step Heckman's model do not consistently estimate the parameter of interest. For example, suppose that unobserved wage determinants (for example ability) are correlated both with education and sector choices, and/or public sector workers are on average more likely to acquire above average education levels due to unobserved factors. In this case standard techniques that control only for selection in the choice of the sector are not able to consistently estimate the effect of working in the public sector and returns to education, and more sophisticated procedures are needed.

3.1 The model and the parameter of interest

Let $\ln w_{iG}, \ln w_{iP}$ to be respectively the (latent) wage that the *i-th* individual earns in the public (G, standing for government) and in the private sector (P):

$$\begin{aligned} \ln w_{Gi} &= \alpha_G + \psi_G S_i + X_i^{'} \beta_G + u_{Gi} \\ \ln w_{Pi} &= \alpha_P + \psi_P S_i + X_i^{'} \beta_P + u_{Pi} \end{aligned} \qquad (1)$$

S is the number of years of schooling and X is a vector of exogenous individual characteristics that influence earnings.

The individual wage may be written as:

$$\ln w_i = D_i \ln w_{Gi} + (1 - D_i) \ln w_{Pi}$$

Using (1) and imposing the restrictive assumption that the parameters are the same in each sub-sample except for the intercept, the wage earned becomes:

$$\ln w_i = \alpha + \psi S_i + X_i^{'} \beta + \delta D_i + \varepsilon_i \qquad (2)$$

where:

$$\alpha = \alpha_P, \psi = \psi_P, \beta = \beta_P, \delta = \alpha_G - \alpha_P \quad \text{and} \quad \varepsilon_i \equiv u_{Pi} + (u_{Gi} - u_{Pi}) D_i$$

Note that (2) is a model with individual-specific effects (random coefficient) for D: due to the presence of u_{Gi}, u_{Pi} the effect of working in the public sector is individual-specific. Nested in this model there is also the specification with homogeneous effects (constant coefficient), for example when $u_{Gi} = u_{Pi}$.

For what concerns investments in education, I assume that optimal individual decisions can be expressed as a linear combination of variables including a random error term:

$$S_i = Z_i^{'} \chi + \tau_i \qquad (3)$$

An individual works in the public sector (D=1) or in the private sector (D=0) as the outcome of an unobserved latent variable D* which can be interpreted as the difference in expected utilities between public and private employment:

$$D_i^* = Z_i^{'} \gamma + v_i$$

Consequently, she chooses to work in the public sector only if this difference is positive (net benefit):

$$D_i = I(Z_i^{'} \gamma + v_i > 0) \qquad (4)$$

where I(A) is an indicator function assuming value 1 whenever A is true.

The probability to be a public employee is influenced both by observable and unobservable factors like individual preferences, attitudes toward the risk, specific monetary or non monetary gains, personal characteristics, family background and tastes for specific job attributes[7].

We allow for arbitrarily correlation between τ and u_P, u_G, v: education is potentially endogenous to sector choices and to wages. Z includes X and a vector H of exogenous variables influencing decisions about sector of employment and investments in education. Z is assumed to be independent of all the error terms and, by writing $Z_i^{'}\chi = X_i^{'}\chi_1 + H_i^{'}\chi_2$, the vector of coefficient χ_2 is different from 0. As specified below, the estimation procedure requires that the equation for sector choice (4) does not include years of schooling (S). As a consequence, the (arbitrary) cross-equations error terms correlation completely captures the association between education and sector choices, and not only the fraction simultaneously determined[8]. I also assume that the disturbance terms contained in the choice equation and in the outcome equations are distributed as a trivariate normal[9]:

$$\begin{pmatrix} u_G \\ u_P \\ v \end{pmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}\begin{pmatrix} \sigma_{GG} & . & \sigma_{GV} \\ & \sigma_{PP} & \sigma_{PV} \\ & & 1 \end{pmatrix}\right] \quad (6)$$

The covariance between the error terms of the two wage equations is not identified since the two wage regimes are not simultaneously observed. Thus, the sector of employment is endogenous to wages: some unobserved characteristics that influence the probability to choose a particular sector of employment could also influence the wage received by the individual once he is employed.

Since the error term in (2) is correlated with the dummy D, the OLS estimator is biased and inconsistent. Note that, even if $u_{Gi} = u_{Pi}$ OLS is biased due to the non zero correlation between u_P, u_G and v.

---

[7] Note that this is a model of "pure choice" since the decision to work in the public sector is not constrained or rationed. In other words, once an individual chooses to work in the public sector (supply side), the (public) employer automatically is willing to hire him (demand side). Of course, this is an unrealistic simplification, especially for the public sector, where recruitment happens through public concourses where the number of applicants is traditionally much higher than the number of available positions. As an alternative, the choice model outlined in (4) may be interpreted as a reduced form for both supply and demand decisions (for a discussion see also Bardasi, 1996).

[8] In this way, the correlation between the error terms of the two equations captures both the part of the sector choices which is explained by the education attainment and the "over and above " residual part which is due to correlation between unobservables driving the two decisions.

[9] Strictly speaking, this is an unnecessary assumption. In fact, it suffices that $v \sim N(0,1)$ and $E(u_j \mid v) = av$ (a linear function), with *j=P, G*. If the error terms are jointly normally distributed this condition follows automatically.

Using a first-order Taylor approximation, the wage differential (treatment effect, TE) between working in the public and the private sector for public sector employees for the *i-th* individual may be written as:

$$TE_i = \frac{w_{Gi} - w_{Pi}}{w_{Pi}} \approx \ln w_{Gi} - \ln w_{Pi} = \delta + (u_{Gi} - u_{Pi})$$

This term has two components: the first one is the coefficient associated to D in (2): $\delta = E(\ln w_{Gi} - \ln w_{Pi} | X_i, S_i) = ATE(X,S)$, where ATE is the (constant across individuals) average gain for a randomly chosen individual with given characteristics. The second component is the individual idiosyncratic effect. The "Average Treatment effect on the Treated (ATT)" may be expressed as:

$$ATT(X,S) \equiv \delta' = E(\ln w_{Gi} - \ln w_{Pi} | X_i, S_i, D_i = 1) = \delta + E(u_{Gi} - u_{Pi} | X_i, S_i, D_i = 1) \qquad (7)$$

which is the average gain in the population plus the average of individual-specific effects among public employees.

Let $u_{Gi} - u_{Pi} = \theta_i$, $u_{Pi} = u_i$, then we may rewrite (2) as follows:

$$\ln w_i = \alpha + \psi S_i + X_i' \beta + \delta' D_i + \omega_i \qquad (8)$$

where $\omega_i = u_i + [\theta_i - E(\theta_i \mid X_i, S_i, D_i = 1)] D_i$

In the discussion of estimation techniques I treat separately the case of a constant effect from the case of individual-specific effects.

Constant public wage premium and "genuine" endogeneity

The composite error in equation (2) contains a term capturing the difference in unobserved factors that influence the wage of public sector workers, with and without working in the public sector. Under the assumption that these factors are the same for each worker ($u_{Gi} = u_{Pi} = u_i$) or mean independent of the decision to be a public employee ($E(u_{Gi} - u_{Pi} | X_i, D_i = 1) = E(u_{Gi} - u_{Pi} | X_i) = 0$)[10], the mean wage differential is equal to δ. In other words, conditional on X, the effect of working in the public sector is assumed to be the same for everyone and independent of sector status. In this case, ATE = ATT and, since there are no individual-specific gains from working in the public sector, δ = δ'. Under the assumption that $u_{Gi} = u_{Pi} = u_i$ (2) becomes:

[10] In other words, ug and up are mean independent of D given X and the difference between the two error terms is unkown or ignored by individuals when they decide to work in the public sector. Thus, their best forecast for this difference is simply zero

$$\ln w_i = \alpha + \psi S_i + X_i^{'}\beta + \delta D_i + u_i \tag{9}$$

However, since D and S are not independent of u_i, the conditional expected value of the error term is different from zero and OLS do not consistently estimate the parameters of the model. Instrumental variable methods applied to a model with two endogenous variables offer a solution to this problem. This paper takes an alternative route. Endogeneity in education is eliminated using standard instrumental variables techniques, whereas correction terms control for endogeneity of sector affiliation by modelling the stochastic dependence between wages and sector choices. To explain the details of this approach, let assume for the moment that the level of education is exogenously assigned. Thus, the model contains one endogenous dummy (D). To derive an estimating equation, just write the conditional expected value of the log wage as[11]:

$$E(\ln w_i \mid D_i, Z_i) = \alpha + \psi S_i + X_i^{'}\beta + \delta D_i + D_i E(u_i \mid D_i = 1, Z_i) + (1 - D_i)E(u_i \mid D_i = 0, Z_i) \tag{10}$$

Heckman (1978, 1979) has shown that, under specific distributional assumptions, selectivity issues when the endogenous variable is a binary treatment can be solved by including the functional form of the conditional expectations for u_i in (10) as an additional variable. Therefore, the equation to be estimated becomes:

$$\ln w_i = \alpha + \psi S_i + X_i^{'}\beta + \delta' D_i + \mu\lambda 1_i + \xi_i \tag{11}$$

where $\mu = \sigma_{PV} = \sigma_{GV} = \sigma$, $\xi_i = u_i - E(u_i \mid D_i, Z_i)$. The conditional mean of the new error term is zero and:

$$\lambda 1_i = D_i \frac{\phi(-Z_i^{'}\gamma)}{\Phi(Z_i^{'}\gamma)} - (1 - D_i)\frac{\phi(-Z_i^{'}\gamma)}{\Phi(-Z_i^{'}\gamma)} \tag{12}$$

the inverse Mill's ratio for the entire sample. Clearly, by including (12) in the wage equation, D can be treated as exogenous. Still, (12) is not known but a consistent estimate for it - $\hat{\lambda 1_i}$ - corresponds to the generalised residual of the first-step probit for the probability to work in the public sector. The wage equation augmented by the correction term (11) can be consistently estimated by OLS in the second step.

If S is not exogenous to the wage, $E(u_i \mid S_i, D_i, Z_i) \neq E(u_i \mid D_i, Z_i)$, so the correction term is different from the generalised probit residual because the (joint) distribution of both the endogenous variables should be taken into account. Therefore, even if the error term has conditional mean zero: $E(\xi_i \mid S_i, D_i, Z_i) = E[(u_i - E(u_i \mid D_i, Z_i)) \mid S_i, D_i, Z_i] = 0$, it is still

[11] I'm leaving implicit the conditional dependence on X and S, since I'm assuming that X is included in Z and S is exogenous.

correlated with S[12] and, as a consequence, (12) does not model properly the stochastic dependence between the endogenous variables and the outcome. The model contains an endogenous variable and OLS do not deliver consistent estimates. It can be shown (see Woolbridge, 2002, pp 567-569) that standard IV techniques offer a straightforward solution to this problem. In fact, (11) can be consistently estimated with a three-stage procedure which combines Heckman methods and instrumental variables techniques: first, the computation of the sample generalized residual $\hat{\lambda 1}_i$ from the probit estimation of (4); second, the application of 2SLS to (11) using [Z, D, $\hat{\lambda 1}$(Z,D)] as instruments for S. Of course, the implementation of this procedure requires an instrumental variable for S. As usual, the test of no selectivity bias (no correlation) is a t-test of μ=0[13].

Individual-specific public wage premium and self selection

In a model with heterogeneous effect the common gain (ATE) is different from the effect of the average individual-specific gain (ATT≠ATE). By writing explicitly (8) and considering for the moment S as exogenous we obtain:

$$\ln w_i = \alpha + \psi S_i + X_i{}' \beta + [\delta + E(u_{Gi} - u_{Pi} \mid Z_i, D_i = 1)]D_i + \{u_i + [\theta_i - E(\theta_i \mid Z_i, D_i = 1)]D_i\} \quad (13)$$

Note that by construction the error term in (13) is zero mean since:

$$E[\theta_i - E(\theta_i \mid Z_i, D_i = 1) \mid Z_i, D_i = 1] = 0 \, .$$

Under the assumption that Z, X and S are independent of u_{Pi}:

$$0 = E(u_{Pi} \mid Z_i) = E(u_{Pi} \mid D_i = 1, Z_i)\, prob(D_i = 1 \mid Z_i) + E(u_{Pi} \mid D_i = 0, Z_i)\, prob(D_i = 0 \mid Z_i)$$

which may be written as:

$$E(u_{Pi} \mid D_i = 1, Z_i) = -E(u_{Pi} \mid D_i = 0, Z_i) \frac{prob(D_i = 0 \mid Z_i)}{prob(D_i = 1 \mid Z_i)}$$

---

[12] In fact the subtraction of E(u | D, Z) from u "cleans" the original error term from the correlation between u and v. What is left in the new error term is the purely random component of u and the part correlated with τ.

[13] Similarly to the standard Heckman model, where OLS statistics are incorrect unless the null of no endogeneity is not rejected, 2SLS statistics should be corrected for the presence of a generated regressor unless $\mu = 0$. Albeit unadjusted standard errors are smaller than the true ones, due to the extreme complexity to derive the correction and to implement it in standard packages like Stata, in what follows I abstract from this problem. However, the comparison of results obtained by estimating the original Heckman model both with the Stata routine (which calculate the true variance covariance matrix) and by just plugging the generalised residual in the second step reveals that standard errors in the two cases are very similar.

From (7)[14], the ATT may be then expressed as:

$$\delta^{'} = \delta + E(u_{Gi} \mid Z_i, D_i = 1) + E(u_{Pi} \mid D_i = 0, Z_i)\frac{prob(D_i = 0 \mid Z_i)}{prob(D_i = 1 \mid Z_i)}$$

Finally, using the fact that Z is independent of all the error terms and the distributional assumptions outlined in (6) the previous expression becomes:

$$\delta^{'} = \delta + (\sigma_{GV} - \sigma_{PV})\frac{\phi(-Z_i\gamma)}{\Phi(Z_i\gamma)} \tag{14}$$

where $(\sigma_{GV}, \sigma_{PV})$ are the cross-equations correlation coefficients between disturbances[15]. The first term in (14) measures the "pure" public sector effect, explained by structural differences between the two sectors that are common to everyone. The second term in (14) capture, on average, the part of the wage differential based on unobservable (by the econometrician: observed by the individual) and individual-specific wage differences for public sector employees. A test of $(\sigma_{GV} = \sigma_{PV})$ is equivalent to test the hypothesis of no selection on unobservable gains and of ATT = ATE. Using (14), (13) may be written as:

$$\ln w_i = \alpha + \psi S_i + X_i^{'}\beta + \delta D_i + (\sigma_{GV} - \sigma_{PV})\frac{\phi(-Z_i^{'}\gamma)}{\Phi(Z_i^{'}\gamma)}D_i + \sigma_{PV}[D_i\frac{\phi(-Z_i^{'}\gamma)}{\Phi(Z_i^{'}\gamma)} - (1-D_i)\frac{\phi(-Z_i^{'}\gamma)}{\Phi(-Z_i^{'}\gamma)}] + \pi_i \tag{15}$$

where π_i is a generic zero mean error term uncorrelated with the regressors.

Since the correction terms contains unknown coefficients, the empirical counterpart of (15) is:

$$\ln w_i = \alpha + \psi S_i + X_i^{'}\beta + \delta D_i + \theta\hat{\lambda}2_i D_i + \mu\hat{\lambda}1_i + \pi_i \tag{16}$$

where hats denote probit estimates and $\theta\hat{\lambda}2_i = (\sigma_{Gv} - \sigma_{Pv})\frac{\phi(-Z_i^{'}\hat{\gamma})}{\Phi(Z_i^{'}\hat{\gamma})}$

The correction term interacted with the dummy captures the individual specific component of the wage premium based on comparative advantages, while the coefficient associated to the generalised probit residual tests for the presence of self selection in the base state of the world (private sector).

Because in my case also education is considered endogenous, the estimation procedure consists of three steps, where in the first step $\hat{\lambda}1_i$ and $\hat{\lambda}2_i$ are computed from a probit for

[14] Here: $\delta' = \delta + E(u_{Gi} - u_{Pi} | Z_i, D_i = 1) = \delta + E(u_{Gi} \mid Z_i, D_i = 1) - E(u_{Pi} \mid Z_i, D_i = 1)$

[15] In a context of switching regressions with endogenous switching these two covariances are the coefficient for the correction terms in the two equations.

(4), and then used as instruments for S in conjunction with Z and D in a 2SLS which involves (16) and (3).

A sample estimate of ATT can be easily obtained as:

$$ATT = \hat{\delta} + \hat{\theta}\frac{\phi(-\bar{Z}_i^{'}\hat{\gamma})}{\Phi(\bar{Z}_i^{'}\hat{\gamma})} \tag{17}$$

where $\bar{Z}_i^{'}$ is the sample mean of Z for public sector workers.

A test for θ = 0 is a test of self-selection on individual unobservable wage differences. Note that, in principle, θ may be positive or negative. The sign of the coefficient may also help to understand which theory – comparative wage advantages or compensating wage differences - is able to explain the behaviour of public sector workers. As discussed above, each individual chooses the sector in which he receives the highest utility (which includes both monetary and non monetary factors). Let assume θ > 0: in this case, public employees are self-selected in the sector where they receive expected monetary gains from participation, and for them the wage premium is somehow larger than for an arbitrary person. In other words, - by assuming that unobserved factors are a proxy for motivation, good matches and other productivity-related wage determinants - they self-select themselves in the sector where they are more productive and where they benefit from a comparative wage advantage with respect to the entire population. Since the allocation of workers is based on productivity-based comparative advantages, it is also efficient. Let instead θ < 0: the unobserved individual productivity of public sector workers is on average lower than their potential productivity in the private sector. Because sector choices are driven by both monetary and non monetary aspects, the amount of non monetary gains they receive from working in the public sector more than compensate wage losses due to low unobserved productivity and comparative disadvantages. In this case, qualitative differences in job attributes – such as wage and employment stability, risk aversion, less stress and competition at the workplace, higher social protections – raise satisfaction at the workplace and more than compensate unobserved (potential) negative wage differences: public employees are willing to pay for these attributes though the wage advantage is lower than for a random individual. Still, since workers are employed in the sector in which they are less productive, the allocation of the workforce across sectors is inefficient.

3.2 The identification strategy

The base specification of the wage equation is very parsimonious and includes (a) a set of controls for gender, age, age squared and dummies for the geographic area of residence (to account for different labour market conditions between north, centre and south of Italy) (b) years of schooling, the dummy for the public sector (c) time dummies for 1998 and 2000 and additional family background variables – years of schooling and occupational status of the parents[16] -, to proxy for the effect of relevant but potentially endogenous individual wage determinants (as the occupation level or the working experience).

Since education attainment and the sector dummy are endogenous, identification usually requires valid sources of exogenous variation. In principle, identification in normally distributed models with self-selection is achieved without exclusion restrictions since the generalised residual is non linear. In other words, the sector choice equation may be estimated by using X and instruments for years of schooling. In models with pure self-selection this strategy has been often criticized since in certain regions of the support the Mill's ratio is linear. Thus, the inclusion of additional variables other than those contained in the wage and education equations may be important to avoid multicollinearity problems and to guarantee identification. In the case of a model with an endogenous dummy rather then pure self-selection, the generalised residual refers to the entire sample and is uncorrelated by construction with the regressors in the selection equation, and collinearity between the correction term and determinants of selection is not a problem (see Vella, 1998). However, to achieve valid identification from an economic point of view, I use instruments both for the sector choice and schooling decisions. This means that at least two variables in Z are not included in X. These are variables that affect the choice of the sector and the level of education but have no direct effect on wages. Compared to a the standard instrumental variable procedure with multiple endogenous variables, the model outlined above is identified under milder assumptions. More precisely, since identification in the choice of the sector is achieved through functional form, the restriction that (at least) one instrumental variable for the sector choice not enters significantly the schooling equation and/or (at least) one instrument for education has no effect on the participation is not requested to identify the model. As usual, all the exogenous variables are included in the two equations for the endogenous processes. The structure of the model also requires the inclusion of D and the correction term(s) (which, in turn, is(are) also function of Z and D) into the education equation. In order to identify the level of education, I follow the strategy of Brunello and

[16] Unemployed, self-employed, white collars, blue collars.

Miniaci (1999), who capture exogenous changes in education through a dummy equal to one for individuals born after 1951, who were 18 years old when, in 1969, a reform liberalised the access to higher education by abolishing the entry exam[17]. However, since returns to education differ across different schooling levels and policy reforms exogenously shift years of schooling only for a specific group of individuals, IV techniques applied in these context are likely to identify local and not general (average) returns. For this reason, additional instruments are desirable. Following Cappellari (2004), I use the geographic birth area as an instrument for schooling, since it is thought to influence human capital accumulation without residual effect on wages, once the impact of the area of residence has been controlled for. To identify the choice of the sector I use two binary variables capturing, respectively, employment of the mother and employment of the father in the public sector. These variables should capture tastes and constraints influencing sector choices. The use of parental background variables as instruments is common in the literature. Still, these variables are often suspected to be positively correlated with unobserved factors in the wage equation and their validity as instruments might be questionable, especially in models without distributional assumptions. I assume that, once the impact of parents' education and occupation has been controlled for, the fact that they worked in public sector it is unlikely to have any significant impact on wages.

Since the model is overidentified, the validity of my identification strategy will be evaluated in the empirical analysis by means of a number of exclusion restrictions' tests.

4. Main Results

For comparative purposes, the first column of Table 2 presents the estimates of (9) by OLS[18]. Results are similar to the existing literature (see for example Comi and Ghinetti, 2002): (marginal) returns to education are on average equal to 4% and those to age (capturing both general and specific on-the-job-formed human capital as well as life cycle wage effects) approximately 5%. The gender wage gap is around 13% and the coefficient associated with the sector dummy (public) shows that public employees earn on average 12% more than private employees (with comparable observable characteristics). Also regional differences

[17] Since this excluded exogenous variable is basically a cohort dummy and age and its square are used as included exogenous variables, identification is obtained by imposing that the same type of information enters wage equations and selection and schooling equations in different way.

[18] From the original set of regressors, I have excluded the set of dummies for parents' occupational status since preliminary estimates showed that the effects of these variables was not significant, once the role of parents' education had been controlled for.

matter: being employed in the north guarantees a wage premium equal to 12% compared to the south, while differences are less pronounced for workers employed in the centre of Italy.

<TABLE 2 AROUND HERE>

The table also reports an informal test on instruments validity. Overall, the data support the exclusion of the set of variables used as instruments (dummy for birth date after 1951, region of birth and dummies for the two parents employed in the public sector) from the wage equation. For what concerns the set of additional controls, only the parents' level of education is statistically significant[19]. If S (educ) and D (pub) were uncorrelated with the error term, then OLS consistently estimate ATT = ATE. However, this is no longer true when education and/or sector choices are not exogenous. The remaining part of the section is organised as follows: for comparative purposes, I first estimate the wage equation corrected for the endogeneity of sector decisions but not for potential endogeneity of education; then I endogenise both processes and estimate the model with the three step procedure discussed in section 3. Table 3 presents the results for the model where the public sector wage premium is assumed to be constant across individuals. First step probit results for the selection equation (reported in column 4) show that the probability to work as a public employee increases when parents have worked in that sector.

<TABLE 3 AROUND HERE>

Other significant determinants of sector choices are the gender (males are less likely to join the public sector, but this maybe reflects the disproportionate number of women among teachers in the sample), the geographic area (of birth and residence) and cohort effects (both the age and being born after 1951). In addition, instruments primarily thought to influence education also affect sector choices. Still, as discussed above, this is not a real problem for the validity of the estimation procedure. Column 3 reports second step OLS estimates of a wage equation augmented by a correction term for endogenous sector choices. Results show that the return to education is equivalent to OLS estimates, while the public wage premium is higher: the estimated wage differential is 24 % and the correction term (lambda1) has a negative coefficient. In other words, the sector dummy is endogenous and, due to unobserved characteristics, workers employed in the public sector earn less in both sectors than the average worker. In order to account for both sources of endogeneity (education and sector decisions), columns 1 and 2 report respectively 3rd and 2nd step results from the application of 2SLS to (11), a wage equation augmented by the correction term estimated in the 1st step.

[19] Results are available upon request from the author. Accor to these results, parental education is not excludable from the wage equation and should not be used to instrument education or sector choices.

Second step estimation of the education equation includes the full set of exogenous variables, including the sector dummy and the correction term(s). Results show that people working in the public sector have on average 3 years of schooling more than private employees and that the education attainment is lower in the north and in the centre. In both cases (constant versus heterogeneous effects) exclusion of instruments is strongly rejected by the data, and instruments for the selection equation (fathpu and motherpu) are not significant in the education equation. Thus, the model is identified also without relying on non-linearities in the functional form of the correction term(s).

For what concerns the estimation of the earning equation (11) in the third step, the overidentification test supports the validity of my identification strategy. A version of Hausman-Wu test for endogeneity of education and sector choices[20] rejects their joint exogeneity with a p-value of 1%. As considered separately, selection in education is negative, although only marginally significant, (maybe because the selection effect is detectable only at relatively high levels of education), while selection in the public sector, captured by the coefficient for lambda1, is negative and significant: OLS underestimate the true treatment effect. The combination of these two endogenous processes also affect the coefficients' estimates. As compared to results in columns 4, important differences emerge: the return to education is approximately equal to 5%, one percentage point more than what standard OLS estimates. This evidence is qualitatively similar to Brunello and Miniaci (1999), who found that returns are underestimated when education is erroneously treated as exogenous due to the sorting of less able individuals in the group with high educational endowment. This may happen because people with higher unobservable productivity drop out school earlier and start working[21]. This behaviour may be rational in Italy, where the (private) economy is mainly composed by traditional and typically low-skill intensity sectors, where specialised positions requiring high levels of cognitive skills acquired through general human capital are quite scarce[22]. On the contrary, the public sector wage premium is equal to 18.5%, smaller than the one estimated in column 4 (24%). The fact that estimates of (11) by OLS are upward biased as compared to results for 2SLS, provides an indirect evidence of positive correlation between

[20] It is implemented including the residuals of the first stage residual of the education equation as an additional regressor in the wage equation. The residual for the sector choice is simply $\lambda 1$, which is already included. Testing whether they are jointly 0 or not is equivalent to test for the presence of endogeneity in S (educ) and D (pub).

[21] As an alternative, liquidity constraints may prevent able pupils with poor family background to make their optimal schooling choice.

[22] Therefore, more productive or more able individuals may have no incentives to acquire additional years of education, because it may be relatively easy to find well-rewarded jobs for them, while being enrolled at school may have relatively high opportunity cost.

unobservable determinants of sector affiliation and schooling decisions: as discussed in section 1, it seems that people with preferences for the public sector are on average more willing to acquire high levels of education. When education is not adequately instrumented the public premium captures both the true effects and the effect of a disproportionately higher number of educated workers among public employees.

As discussed in section 3, this model has the disadvantage that, since unobservable wage determinants are assumed to be equal across the two sectors, the coefficient for the endogenous dummy estimate both the average premium for a random individual and for those actually working in the public sector.

<TABLE 4 AROUND HERE>

Table 4 reports the estimates of the more flexible specification (16), which allows for individual-specific returns to sector choices and self selection. Column 3 presents the results of the model estimated treating education as exogenous. The wage premium for an individual randomly selected from the population (ATE) is about 18%. Since the coefficient for lambda1 is not significantly different from zero, workers in the base state of the word (private sector) are not self-selected on the basis of unobserved characteristic: their earning potential is not different from that of the average worker. On the opposite, there is significant negative self-selection in the public sector, as revealed by the coefficient for lambda2*pub (see Bardasi, 1996 for similar results). Since the difference between earnings potential in the two sectors for public employee is negative, their unobserved characteristics could allow them to earn more in the private sector. This implies that the allocation of skills is inefficient, and workers self-select themselves into the sector where their productivity is less rewarded. As a consequence, the choice to work in the public sector seems to be based more on (unobservable) non monetary factors compensating monetary losses than on comparative wage advantages. As a consequence of negative self-selection, the coefficient associated with the dummy for sector affiliation (which represents the treatment effect in the population) overestimate the true average wage differential for public workers. In fact the Average Treatment for the Treated (ATT), obtained by adding the interaction term (lambda2*pub) evaluated at the average sample characteristics to the sector dummy coefficient as in (18), is still positive but equal to 10%.

Next, columns 1 and 2 presents the results obtained when education and sector decisions are considered endogenous. Results are consistent to those reported in table 3 and qualitatively similar: the return to education increases at the level of 5% and the public wage premium decrease from 18% to 14.5%. The coefficient associated to the interaction term (lambda2*pub), which captures the individual-specific component of the wage premium, is again negative and significant.

Overall, this evidence may be interpreted in the following way: workers who acquire more education are on average those with low ability on the labour market and, also, those who are more likely to self-select themselves into the public sector. For them, the average wage differential (ATT) is approximately equal to 8.5%.

5. Conclusions

The aim of this paper was to investigate the effect of endogenous sector choices and schooling decisions on the wage premium received by Italian public employees during the 90's. Existing literature on public/private wage differentials for Italy dealt with the first selectivity source only, typically estimating earning functions with standard two-stage Heckman methods. However, according to the existing empirical literature (also for Italy), education decisions are likely to be correlated with unobservable wage determinants (and maybe also with sector choices). If not adequately taken into account, this additional source of endogeneity is likely to produce biased and inconsistent estimates of the whole set of parameters, including the wage premium. Two alternative specifications were discussed: in the first one the public sector wage effect is homogeneous in the population and captured by a public sector dummy coefficient; in the second one the coefficient is random and the effect contains an individual-specific component given by the difference between unobservable wage potential in the two sectors. The parameter of interest is the ATT (Average Treatment effect on the Treated), i.e. the wage effect of working in the public sector for public sector employees. In order to evaluate how educational choices affects the wage premium, in the empirical strategy I compared the results from two specifications: first, a traditional two-step Heckman procedure - in which only endogeneity of sector choices is accounted for –; then, a 3 stage procedure - which combines the Heckman selection model (to correct for endogenous sector choices) with 2SLS techniques (to eliminate the effect of endogenous schooling decisions). The data set was obtained by pooling information from 1995, 1998 and 2000 Bank of Italy Households' Surveys. A common result is that, if not adequately treated, endogeneity of education lowers returns to education - due to the (weak) sorting of less able individuals in the group with high schooling attainment -, and produces an upward bias in the estimate of the public wage premium in both the specifications. Albeit a direct evidence in this sense was lacking, it is reasonable to assume that this result largely depends on a positive correlation between (unobserved) determinants of schooling decisions and sector choices: according to this interpretation, workers with high school levels (and with ability on the labour market lower than the average) are more likely to show preferences for public employment, which is consistent with the idea that public workers strategically acquire high levels of education in order to be recruited by the public sector. These results also suggest that existing estimates of the public/private wage differential, obtained by controlling only for selection into the two sectors, are higher than the true one.

By concentrating on the results from the more general 3 step procedure, under the more restrictive assumption of a constant return to public employment, the estimate of the wage premium in the population (which, in this case, coincides with the effect for public employees) is 18%, higher than OLS estimates since public workers are endowed with an (unobserved) wage potential below the average in both sectors.

Under the more flexible specification with individual-specific returns to public employment, which relaxes the restrictive assumption of equal unobserved wage determinants across sectors, the wage differential for a randomly chosen worker is equal to 15%. For what concerns public employees, they are negatively self selected: in fact, they have a lower unobservable productivity in the sector in which they actually work and, as a consequence, their wage premium (ATT) is still positive (8.5%) but lower than for a random individual. In other words, public workers are employed in the sector in which their earning potential is lower, and, as a results, they are willing to renounce to the comparative wage advantage offered by the private sector for non-monetary job attributes attached to public positions that more than compensate potential monetary losses.

These results suggest a number of considerations about the relative efficiency of public sector retaining, recruiting and pay policies. A well-known fact in Italy is that the wage structure is compressed, especially in the public sector, which pays a lot relative to the private sector for low-skilled positions and less for the high skills.

From the individual perspective, able workers with good labour market opportunities not requiring high levels of cognitive skills have an incentive to drop out school and to start working in the private sector. On the contrary, less able individuals with worst outside options may have an incentive to acquire higher levels of education (public education in Italy is relatively cheap and the alternative is being unemployed), partly to increase productivity and, possibly, to apply for the public sector, which pays more than the private sector and where holding high levels of education is interpreted as a signal of ability in the recruitment procedure. On the whole, public employees would be more motivated, productive and better matched in the private sector. Still, they apply for a position in the public employment because it pays a ceteris paribus premium relative to the private sector and guarantees higher levels of valuable non-wage job attributes, like stability and flexibility.

From the public sector perspective, this situation creates inefficiencies: on the one hand, recruitment methodologies based on schooling performance as a signal of high ability are not effective and are likely to select individuals with low ability; on the other hand, due to negative self-selection into the public sector, non-monetary job attributes are likely to attract individuals with comparable low productivity. As a result, the allocation of skills across sectors is inefficient and public employees are less productive than what they would be if employed in the private sector. The distortions created by this mechanism are not negligible: the private sector attracts able and productive workers. The public sector, by paying a wage premium and offering non-wage benefits, such as a good working environment, job security, flexibility, lower levels of effort, etc. (valuable to its employees, who are traditionally risk-adverse), selects a disproportionately higher number workers with low ability and (relative) low productivity. In other words, on the one hand public employees are well matched with public sector positions for what concerns working conditions and non-wage attributes; on the other hand, are badly matched for what concerns productivity-related measures. Albeit this situation is may be optimal from the point of view of public employees, efficiency and equity considerations suggest that it might not be totally desirable from a social perspective. In this context, reforms of wage determination in the public sector aimed at increasing productivity and efficiency levels are desirable but probably do not suffice, and new rules and procedures to recruit and retain workers may be desirable as well.

Of course, results are conditional to implicit and explicit model assumptions. For example, constraining earnings profiles to be equal across sectors (except for the intercept) is a simplification which may hide differences between personal characteristics' returns across sectors. In addition, regional differences are a dimension which probably deserve further investigation, similarly to the analysis of the distribution of the public premium across the skill distribution. In this context, the estimation of a more general specification, which relaxes some of the assumptions about the distribution of the errors and the structure of the model is probably needed and is left for future research.

REFERENCES

Adamchick, V., and Bedi, B., 2000, Wage differentials between the public and the private sectors: Evidence from an economy in transitions, Labour Economics, 7, 203-224.

Alesina, A., Danninger, S. and Rostagno, M., 2001, Redistribution Through Public Employment: The Case of Italy, IMF Staff Papers, vol. 48, n. 3.

Bardasi, E., 1996, I differenziali salariali tra pubblico e privato: un'analisi microeconometrica, Lavoro e Relazioni Industriali, 3.

Belman, D. and Heywood, J. S., 1989, Government wage differentials: a sample selection approach", Applied Economics, 21.

Brunello, G. and Dustmann, C., 1997, Le retribuzioni nel settore pubblico e privato in Italia: un paragone basato su dati microeconomici, in *Rapporto Aran sulle retribuzioni*, ed. by Dell'Aringa, Franco Angeli: Roma.

Brunello, G. and Miniaci, R., 1999, The economic returns to schooling for Italian men: an evaluation based on instrumental variable, Labour Economics, 6, 509-19

Brunello, G. and Rizzi D., 1993, I differenziali retributivi nei settori pubblico e privato in Italia: un'analisi cross-section, Politica Economica, 3.

Cannari L., Pellegrini, G. and Sestito P., 1989, Redditi da lavoro dipendente: un'analisi in termini di capitale umano, Temi di discussione Banca d'Italia # 124.

Cappellari, L., 2002, Earnings dynamics and uncertainty in Italy: how do they differ between the private and public sector?, Labour Economics, 9(4).

Cappellari, L., 2004, Earnings Mobility Among Italian Low Paid Workers, IZA Discussion Paper, n. 1068.

Colussi, A., 1997, Il tasso di rendimento dell'istruzione in Italia: un'analisi cross-section, Politica Economica, vol. XIII, n. 2.

Comi, S. and Ghinetti, P., 2002, La distribuzione dei salari in Italia: un confronto tra pubblico e privato, Rivista Internazionale di Scienze Sociali, vol. CX.

Disney, R. and Gosling, A., 1998, Does it pay to work in the public sector?, Fiscal Studies, 19(4), 347-374.

Dustmann C. and Van Soest, 1998, Public and Private sector wages of male workers in Germany, European Economic Review, 42, 1417-1441.

Hartog, J and Oosterbeek, H, 1993, Public and private sector wages in the Netherland, European Economic Review, 37.

Heckman, J. J., 1978, Dummy endogenous variables in simultaneous equation systems, Econometrica, 46.

Heckman, J. J., 1979, Sample selection as a specification error, Econometrica, 47.

Heckman, J. J. and Robb, R. Jr, 1985, Alternative methods for evaluating the impact of interventions, in *Longitudynal analysis of labour market data*, ed. By Heckman and Singer, New York: Cambridge Univ. Press.

Lucifora, C., 1999, Rules Vs bargaining: pay determination in the Italian public sector, in Public sector pay determination in the European Union, MacMillan.

Smith, S., 1977, Government wage differentials, Journal of Urban Economics, 4.

Vella, F., 1998, Estimating models with sample selection bias: a survey, Journal of Human Resources, 33.

Willis, 1986, Public Sector labour markets, Handbook of Labour Economics.

Wooldridge, J. M., 2002, Econometric Analysis of Cross-Section and Panel Data, MIT press.

TABLES

Table 1. Summary statistics

		Whole sample		Private sector		Public sector	
Variable	Description	Mean	Stdv	Mean	Stdv	Mean	Stdv
lnhwage	ln(hourly wage)	2.455	0.449	2.365	0.447	2.659	0.383
educ	years of schooling	11.084	3.851	10.328	3.602	12.803	3.846
pub	public sector = 1	0.306	0.461				
age	Age (years)	39.182	10.752	37.461	10.988	43.088	9.055
age2	age squared	1650.82	852.19	1524.07	850.32	1938.59	783.79
male	Male = 1	0.598	0.490	0.642	0.480	0.498	0.500
Regio resid north	residence in the north = 1	0.484	0.500	0.532	0.499	0.373	0.484
Regio resid centre	residence in the centre = 1	0.217	0.412	0.215	0.411	0.221	0.415
regre3	residence in the south = 1	0.300	0.458	0.253	0.435	0.406	0.491
d95	1995 = 1	0.347	0.476	0.312	0.463	0.427	0.495
d98	1998 = 1	0.313	0.464	0.324	0.468	0.290	0.454
d00	2000 = 1	0.340	0.474	0.365	0.481	0.283	0.451
edfath	Father education	4.489	4.498	3.961	4.288	5.687	4.726
edmoth	Mother education	4.027	4.125	3.656	4.084	4.871	4.096
fathpu	father public employee	0.142	0.349	0.106	0.307	0.226	0.418
mothpu	mother public employee	0.052	0.223	0.039	0.194	0.083	0.275
1951	year of birth after 1951	0.730	0.444	0.769	0.421	0.641	0.480
Regio birth north	Birth in the north = 1	0.422	0.494	0.469	0.499	0.313	0.464
Regio birth centre	Birth in the centre = 1	0.205	0.404	0.207	0.405	0.202	0.401
regbi3	Birth in the south = 1	0.373	0.484	0.324	0.468	0.485	0.500
N. obs		*16568*		*11502*		*5066*	

Table 2. Wage equation with constant public sector premium: OLS results

Dep var: ln(hourly wages)	Coef.	t-stat
Education	0.040	(46.82)
Public	0.12	(18.50)
Age	0.049	(22.82)
Age2	-0.0004	(16.31)
Male	0.129	(21.86)
Regio resid1	0.123	(17.22)
Regio resid2	0.046	(5.58)
Intercept	0.58	(13.59)
Test: Exclusion of instruments from the wage equation	Pr>F = 0.536	
R2		0.34
# observations		16564

Note: Pooled data for 1995, 98 and 2000. t-statistics (robust to heterosk. and autocorr.) in parenthesis. Additional controls included in the regression: time dummies for 1998 and 2000, parents' education level. Excluded categories: year 1995, South, father unemployed. Test statistics are from Wald test for excluded instruments (H0: d_1951 = regio birth North = regio birth Centre = fathpub = mothpub = 0).

Table 3: Wage equation with constant public sector premium: two- and three-step procedures

	education and sector endogenous				Only sector endogenous			
	3rd step: wage equation		2nd step: education equation		2nd step: wage equation		1st step: sector equation	
	(1)		(2)		(3)		(4)	
Variables	Depvar: ln(hourly wages)		Depvar: educ		Depvar: ln(hourly wages)		Depvar: Public	
Intercept	0.504	(6.59)	9.509	(28.46)	0.608	(14.03)	-3.463	(21.3)
Educ	0.0513	(8.02)			0.0404	(46.69)		
Pub (ATE = ATT)	0.185	(3.70)	5.135	(10.78)	0.227	(5.16)		
Age	0.0456	(18.16)	0.0149	(0.65)	0.045	(18.26)	0.137	(15.63)
Age2	-0.0004	(13.28)	-0.0014	(4.86)	-0.0004	(13.99)	-0.001	(10.53)
Male	0.148	(16.23)	-0.13	(1.54)	0.145	(16.51)	-0.471	(20.78)
Regio resid1	0.1381	(12.02)	-0.591	(4.76)	0.143	(12.81)	-0.436	(10.22)
Regio resid2	0.055	(5.46)	-0.458	(2.95)	0.058	(5.82)	-0.184	(3..18)
Lambda1	-0.0547	(2.02)	-1.649	(5.81)	-0.064	(2.44)		
Instruments:								
Mothpub			-0.042	(0.34)			0.235	(4.57)
Fathpub			0.0535	(0.56)			0.285	(8.92)
1951			-0.4156	(3.41)			0.216	(5.09)
Regio birth1			1.541	(14.97)			-0.149	(3.55)
Regio birth2			0.9531	(6.31)			-0.155	(2.71)
Test 1: Exclusion of instruments from the sector equation							*0.000*	
Test 2: Exclusion of instruments from the education equation			*0.000*					

Table 3. – continued -

Test3: exogeneity of education	*0.083*			
Test 4: exogeneity of education and sector choice	*0.012*			
Test 5. Overidentific. Restrictions	*0.908*			
Centred R2	*0.34*		*0.37*	
R2		*0.235*		
Log-likelihood				-*8722.04*

Notes: Pooled data for 1995, 98 and 2000. n° of obs.: 16564. Estimates: in column (4) are obtained by probit; in column (3) by OLS; in columns (1) and (2) by 2SLS. t-(or z-)statistics (robust to heterosk. and autocorr.) in parenthesis. Additional controls included in the regression: time dummies for 1998 and 2000, parent's education. Excluded categories: year 1995, South. Test 1-2 give statistics from Wald test of hypoteses, p-values reported in italics. Test 1: validity of instruments in the probit for the sector choice (H0: d_1951=regio birth North=regio birth Centre=fathpub=mothpub=0). Test 2: validity of instruments in the education equation (H0: d_1951=regio birth North=regio birth Centre=fathpub=mothpub=0). Test 3: significativity of the residual from education equation (H0: residual education equation=0). Test 4: joint significativity of residuals from the sector and education equation (H0: residual education equation=generalised residual of the sector equation (lambda1)=0). Test 5: gives Hansen j-statistic for the overdentification of all instruments in a 2SLS prodedure. §: obtained from (17) in section 3.

Table 4. Wage equation with individual-specific public sector premium: two- and three-step procedures

	education and sector endogenous				Only sector endogenous			
	3^{rd} step: wage equation		2^{nd} step: education equation		2nd step: wage equation with correction terms		1st step: sector equation	
	(1)		(2)		(3)		(4)	
Variables:	Depvar: ln(hourly wages)		Depvar: educ		Depvar: ln(hourly wages)		Depvar: pub	
Intercept	0.495	(6.55)	8.83	(24.33)	0.570	(12.65)	-3.4501	(20.54)
Educ	0.049	(7.33)			0.040	(46.54)		
Pub (ATE)	0.148	(2.92)	3.749	(6.58)	0.173	(3.65)		
Age	0.049	(17.18)	0.0804	(2.95)	0.049	(17.66)	0.137	(15.63)
Age2	-0.004	(12.90)	-0.0019	(6.23)	-0.0004	(14.12)	-0.0012	(10.53)
Male	0.137	(13.36)	-0.377	(3.74)	0.133	(13.91)	-0.471	(20.78)
Regio resid north	0.127	(10.48)	-0.817	(6.12)	0.129	(10.69)	-0.437	(10.22)
Regio resid Centre	0.05	(4.80)	-0.541	(3.42)	0.051	(4.93)	-0.184	(3.18)
Lambda1 (σ_{PV})	0.013	(0.36)	-0.191	(0.43)	0.017	(0.48)		
Lambda2*pub (σ_{GV} - σ_{PV})	-0.065	(2.52)	-0.974	(4.03)	-0.076	(3.15)		
Instruments:								
Mothpub			0.0771	(0.59)			0.236	(4.57)
Fathpub			0.213	(2.08)			0.285	(8.92)
1951			-0.291	(2.33)			0.216	(5.09)
Regio birth north			1.462	(13.99)			-0.149	(3.55)
Regio birth centre			0.861	(5.66)			-0.155	(2.71)
ATT[§]	0.087				0.101			
Test 1: Exclusion of instruments from the sector equation							*0.000*	
Test 2: Exclusion of instruments from the education equation			*0.000*					

Table 4. – continued -

Test 3: Exogeneity of education	*0.098*			
Test 4: exogeneity of education and sector choice.	*0.002*			
Test 5: overidentifying restrictions	*0.853*			
Centred R2	0.36			
R2		0.235	0.37	
Log likelihood				-8403.8

Notes: Pooled data for 1995, 98 and 2000. n° of obs.: 16564. Estimates: in column (4) are obtained by probit; in column (3) by OLS; in columns (1) and (2) by 2SLS. t-(or z-)statistics (robust to heterosk. and autocorr.) in parenthesis. Additional controls included in the regression: time dummies for 1998 and 2000, parent's education. Excluded categories: year 1995, South. Test 1-4 give statistics from Wald test of hypoteses, p-values reported in italics. Test 1: validity of instruments in the probit for the sector choice (H0: d_1951=regio birth North=regio birth Centre=fathpub=mothpub=0). Test 2: validity of instruments in the education equation (H0: d_1951=regio birth North=regio birth Centre=fathpub=mothpub=0). Test 3: significativity of the residual from education equation, coeff.=-0.009 t-stat=1.30 (H0: residual education equation=0). Test 4: joint significativity of residuals from education and sector choice equation (H0: residual education equation=generalised probit residual for sector choice (lambda1)=0). Test 5 gives Hansen j-statistic for the overdentification of all instruments in a 2SLS prodedure. §: obtained from (17) in section 3.