Consistent Estimates of the Public/Private Wage Gap

Domenico Depalo & Santiago Pereda-Fernández *

Abstract

Existing estimates of the public/private wage gap allow for possible sorting of individuals into one sector, but they rely on parametric assumptions that may introduce substantial bias in the parameter of interest. Solutions are semi and nonparametric approaches. For Italy, the latter methods yield a gap of approximately 20%, whereas the bias from parametric assumptions is as large as 10%.

JEL classification: J31, J38, C14

Keywords: public/private wage gap, nonparametric regression, sample selection

*We thank two anonymous referees for their constructive comments. All errors are our responsibilities. Replication files and additional results will be available at the webpage: http://sites.google.com/site/domdepalo/ The views expressed in this paper are those of the authors and do not imply any responsibility of the Bank of Italy. Corresponding addresses: Domenico Depalo, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Rome (Italy), Tel.: +39-06-4792 5989, e-mail: domenico.depalo@bancaditalia.it Santiago Pereda-Fernández, Banca d'Italia, Economics and Statistics Department, Via Nazionale, 91 - 00184 Rome (Italy), Tel.: +39-06-4792 3892, e-mail: santiago.pereda@bancaditalia.it

1 Introduction

Representing about one quarter of total public expenditure (source: AMECO), the wage bill is one of the most important tools at the disposal of governments to contain such expenditure. The Great Recession has renewed the interest of policy makers in the size of the public/private wage gap (PPWG), for which it is necessary to provide a reliable estimate. This is the goal of this paper.

Accounting for sorting into sector is crucial to obtain consistent estimates of the PPWG, since otherwise there would be a bias proportional to the severity of selection. Much of the existing literature in this field exploits a switching endogenous dummy model (Lausěv, 2014), imposing parametric assumptions and a constant treatment effect (Imbens and Rubin, 1997).

However, and this is the main contribution of this paper, even under constant treatment effect a slight departure from any maintained parametric assumption would lead to inconsistent estimates of the PPWG. To the best of our knowledge, this issue has not been noticed in the existing literature (see the review in Lausěv, 2014 and the papers appeared thereafter). Therefore, we assess the sensitivity of the estimates of the PPWG by using a variety of methods that subsequently relax parametric and functional form assumptions.

Another contribution of this paper is to propose a simple rule of thumb to gauge the credibility of the constant treatment effect restriction, that simplifies identification issues arising with heterogeneous effects considerably (Vytlacil, 2002). This matter is explicitly addressed by Depalo (2018) in the estimation of the PPWG.

As for the empirical contribution of the paper, we focus on the Italian PPWG. We find that imposing normality on the unobservable characteristics has a major impact on the magnitude of the estimated gap, and using either semi or nonparametric methods yield significantly different estimates. Parametric methods estimate a gap as large as 30%, compared to 20% with either the semi or nonparametric approaches. Moreover, we find evidence of heterogeneous effects.

2 The Model

Following the existing literature (Lausěv, 2014), we consider:

$$d_i = \mathbf{1}\left(\delta\left(z_i\right) + \nu_i > 0\right) \qquad\qquad i = 1, \ldots, N \tag{1}$$

$$y^*_{di} = g_d\left(x_i\right) + \varepsilon_{di} \tag{2}$$

$$y_i = d_i y^*_{1i} + \left(1 - d_i\right) y^*_{0i} \tag{3}$$

where $d_i = 1$ for public sector, y^*_{1i} and y^*_{0i} are the potential wages in the public and the private sectors, and y_i is the observed wage; x_i are the covariates, z_{1i} is an exogenous and relevant instrument, and $z_i = [x_i, z_{1i}]$; ε_{0i}, ε_{1i}, and ν_i are unobserved error terms that are potentially correlated, and $\mathbf{1}(\cdot)$ is an indicator function.

The parameter of interest is the PPWG, defined as the average differential in earnings between public and private sectors across the whole population, i.e. $\mathbb{E}\left[y^*_{1i} - y^*_{0i}\right]$. Public sector workers may conform a non-random sample of the population, e.g. because they may be more risk averse (Lausěv, 2014). Consequently, $\mathbb{E}\left[\varepsilon_{ji} | d_i = j\right] \neq 0$ for $j = \{0, 1\}$, so OLS estimates of the PPWG are inconsistent (Maddala, 1983).

Various corrections for the selection bias have been proposed. Early works assumed a parametric distribution for the error terms (parametric estimators), namely multivariate normal, as well as linearity of the regressors: $\delta\left(z_i\right) = z_i'\gamma$, $g_d\left(x_i\right) = x_i'\beta_d$.[1] Under this assumption, a consistent estimator of β may be computed in two steps (Heckman, 1979): first, run a probit of d_i on z_i, obtaining $\hat{\gamma}$; second, regress y_i on $d_i x_i$, $\left(1 - d_i\right) x_i$, $d_i \frac{\phi\left(z_i'\hat{\gamma}\right)}{\Phi\left(z_i'\hat{\gamma}\right)}$, and $\left(1 - d_i\right) \frac{\phi\left(z_i'\hat{\gamma}\right)}{1-\Phi\left(z_i'\hat{\gamma}\right)}$, where $\phi(\cdot)$ is the normal density function, and $\Phi(\cdot)$ its cumulative distribution. These components identify the PPWG. Alternatively, since the system is fully parametric, one can use maximum likelihood estimation (MLE).

If the imposed parametric assumptions are correct, both estimators are consistent, although MLE is asymptotically more efficient. A drawback of parametric estimators is that they are incon-

[1] In this note we assume the existence of an exclusion restriction. However, this is not strictly required in the fully parametric model.

sistent under even minimal departures from their functional assumptions. Arabmazar and Schmidt (1982) show that the bias from normality can be substantial, particularly when the disturbance variance is unknown.[2] In fact, the bias of the MLE can be larger than the bias of the uncorrected sample mean.

Therefore, it is essential to relax the parametric assumptions. A natural way to do this is to use a semiparametric estimator, such as Newey (1999, 2009). This estimator has the advantage of not requiring to specify the parametric distribution, although it requires the wage equation to be linear. More specifically, it assumes that, conditional on selection, the mean of the disturbance in the wage of each sector depends only on an unknown index $\delta_i = \delta(z_i, \gamma)$. Hence, $\mathbb{E}\left[y_i|z_i, d_i = j\right] = x_i'\beta_d + h\left(\delta_i\right)$, with $h\left(\delta_i\right) = \mathbb{E}\left[\varepsilon_i|z_i, d_i = j\right]$. The estimator is implemented in two steps: first, regress d_i on a polynomial of z_i, obtaining the propensity score $\hat{p}_i$; then, regress y_i on $d_i x_i$, $\left(1 - d_i\right) x_i$, $\left\{d_i \hat{p}_i^k\right\}_{k=1}^K$, and $\left\{\left(1 - d_i\right)\left(1 - \hat{p}_i\right)^k\right\}_{k=1}^K$, getting an estimate of the PPWG.[3]

A further relaxation of the assumptions is to allow for non-linearities in the wage equation, *e.g.* a functional form $g_d\left(x\right)$, and then estimate both the wage *and* the selection equations with a flexible estimator. Formally, $\mathbb{E}\left[y_i|z_i, d_i = j\right] = g_j\left(x_i\right) + h\left(\delta_i\right)$. Das et al. (2003) show the identification of $g_d\left(x_i\right)$, and proposed a two-step estimator: first, nonparametrically regress d_i on z_i to get the propensity score; then regress y_i on $\left\{d_i q_{1k}\left(x_i, \hat{p}_i\right)\right\}_{k=1}^K$, and $\left\{\left(1 - d_i\right) q_{0k}\left(x_i, 1 - \hat{p}_i\right)\right\}_{k=1}^K$, where q_{jk} is an approximation function for $j = \{0, 1\}$. In our case, we set $q_{j,k}\left(x, p\right)$ to include all quadratic terms and interactions of the covariates, on top of a polynomial of the propensity score.

The estimators proposed by Newey (2009) and Das et al. (2003) estimators are very flexible and based on weaker assumptions, making them more robust than those relying on parametric assumptions. On the other hand, aside from considerations on efficiency, they require an exclusion restriction; they require a trimming function to guarantee that the propensity score estimates are strictly inside the unit interval; they require the choice of the polynomial order of the series (K), e.g., through cross-validation.

[2]For illustrative purposes, we provide formal intuition on the size of the bias of the Heckman two-step estimator in Appendix A.

[3]Newey (2009) considers power series and splines for the second stage (denoted by $\{\cdot\}_{k=1}^K$), and some semiparametric methods to estimate the first stage. For a more transparent comparison to Das et al. (2003), we use power series in the wage equation and the linear probability model in the selection equation.

Our list of estimators is not exhaustive, but the estimators presented here are either common in the literature or relatively straightforward to interpret and implement.[4] Also, they satisfy the conditions for nonparametric identification of counterfactual outcomes (Heckman and Vytlacil, 2007).

3 Data and Results

We use data from the Survey on Household Income & Wealth (SHIW), which is conducted every two years by the Bank of Italy on a sample representative of the Italian population. The data contain information about a wide range of personal characteristics like age, gender, marital status, educational level, and region of residency, and occupational characteristics, like sector of economic activity, occupational level, firm size, professional condition, part-time status, number of months worked in the year, and average number of hours worked in a week. Crucially for this paper, wages net of income and payroll taxes are reported. These data have been largely employed to study the PPWG in Italy (Brunello and Dustmann, 1997; Lucifora and Meurs, 2006; Depalo and Giordano, 2011; Depalo, 2018).

The period covered by the analysis is 1998-2012. We focus on men, in the age range 25-64 (Dustmann and van Soest, 1998); public sector employees include those whose sector of activity is "public administration, defense, education, health and other public services" (Lucifora and Meurs, 2006). The dependent variable is log hourly wage. Following the existing literature, the set of covariates includes age, and dummies for marital status, educational achievement, job position, full-time employment, geographical areas, years. This leaves us with a sample size of $20,456$ observations. In our sample, public sector employees unconditionally earn on average 20% more than their private sector counterparts, are 6 years older and better educated (Table 1). The rest of the paper investigates whether the wage gap persists after taking into account differences in these observable characteristics.

The performance of all the estimators that we consider relies on the validity of the instrument, which in this application is the sector of the father. This instrument has a long tradition in the field,

[4] For a review on sample selection estimators, see Vella (1998).

although there might be reasons to believe that parents characteristics could affect wages indirectly (Card, 1999), which would threaten the validity of the instrument. Therefore, we formally test the joint null hypotheses of exclusion restriction and monotonicity (Mourifié and Wan, 2017).[5] After controlling for educational attainment and years, the joint null hypothesis is not rejected at standard confidence levels (Table 2).[6] The sector of the father is also relevant in shaping individual preferences in favor of the public sector (Depalo and Giordano, 2011), because preference for the public sector exhibits an inter-temporal persistence from fathers to children (Brunello and Dustmann, 1997; Dustmann and van Soest, 1998). In our data, 1 in 4 public sector employees had a father working in the public sector, as opposed to the 1 in 10 for private sector employees.

Table 3 shows the estimates of the PPWG. IV and Heckman (two-step and ML) are the standard in the literature, whereas Newey (2009) and Das et al. (2003) are our proposed approaches. For completeness, OLS is reported.[7] OLS estimates a gap of less than 6%, significantly smaller than either the IV or Heckman estimates, all of which range between 16 and 32%.

The first important finding is that the minimum and the maximum values of the gap are those of the parametric estimators. Under normality, both estimators are consistent, but in this application the estimates are markedly different. Since one of them is based on moment conditions, and the other one on the whole distribution of the error terms, they may be differently biased when the parametric assumption is wrong. To the best of our knowledge, no such claim has been made in the existing literature in this field.[8] Incidentally, this setup suggests a (quasi) Hausman test (Nijman and Verbeek, 1992), unless one relaxes the parametric assumptions as we do next.

---

[5] The test takes the form of two inequalities that are necessary to identify a Local Average Treatment Effect (Imbens and Angrist, 1994): $P[y, D = 1|Z = 0] \leq P[y, D = 1|Z = 1]$ and $P[y, D = 0|Z = 1] \leq P[y, D = 0|Z = 0]$. If any of the two inequalities is violated, the validity of the instrument is falsified. In the latter case, the test is not informative about which of the two assumptions fails.

[6] As a further check in favor of the validity of the instrument, we estimate the model with IV using a more parsimonious and plausibly exogenous set of controls and then add them progressively: under the maintained assumption of constant treatment effect, finding stable results would be additional evidence in favor of the validity of the instrument and the method(s) in general; if there is instability in the estimated effect, nothing can be said. Controlling only for age education and year, the wage differential is slightly above 30% (Table 6); as soon as we add the job position, it marginally decreases to 28%, where it remains no matter what variables we add.

[7] Standard errors for the OLS, IV, and Heckman estimators were calculated using the asymptotic formula; those for the Newey (2009) and Das et al. (2003) estimators were bootstrapped.

[8] The difference between the two estimators could be due to the relevance of the exclusion restriction (Puhani, 2000). In our data, the F-statistic is equal to 257.6 , a very large value (Bound et al., 1995). Therefore, we argue that the distribution assumptions play a major role.

For both the semi and nonparametric estimators, we set the polynomial order of the series to $K = 1, \ldots, 5$: we view this choice as a fair compromise between flexibility (to control for non-normality) and efficiency (in terms of degrees of freedom).[9] Because the estimates of the PPWG remain stable as K changes, we did not select K using cross-validation.

The semiparametric estimator that relaxes normality but retains the linear wage equation yields an estimated PPWG close to 20%, regardless of the order of the polynomial of the propensity score. The second important empirical finding of this paper is that this gap is substantially different from any of the parametrically estimated gaps, suggesting that the normality assumption is too strong and yields biased estimates.

As with the semiparametric estimator, the estimates of the nonparametric estimator are largely constant over K, with an estimated gap of 20%. The final important empirical finding is that this figure is almost identical to the one found with the linear specification of the semiparametric approach, meaning that non-linearities in the wage equation have at most a tiny effect on the estimated gap. Thus, the bias can be almost entirely attributed to the normality assumption.

Regarding the coefficients of the covariates used in the estimation, the results are consistent with those existing in the literature (Table 4). In particular, we find that earnings have an increasing profile with age, that we take as a proxy for seniority and labor market experience. Moreover, increasing the level of education is associated with higher wages: relative to those with a low level of education, those who completed secondary education have an extra 10% earnings, whereas those with a college degree earn about 20% more. Similarly, wages for managerial positions are about 25% higher than for white-collar positions. Lastly, wages are 10% higher for married workers.

3.1 Constant vs Heterogeneous treatment effect

To conclude the analysis, we assess the plausibility of the constant treatment effect assumption. We suggest two (complementary) approaches. If there exist more than one valid instrument, one can simply compare the estimates with each of them. Under the assumption of constant treatment effect, the differences in the PPWG should be negligible.

[9] Note that the top and bottom 1% of the sample is trimmed in the estimation of the semi and nonparametric estimators to guarantee that the propensity score estimates are strictly inside the unit interval.

To this aim, we use the sector of the mother as an instrument. With some exceptions (particularly at low educational attainment), its validity is not falsified (Table 5). However, it is less relevant to explain sector choice than the sector of the father, as only 5% of the private sector employees had a mother working in the public sector, as opposed to 9% of the public sector employees (Table 1); the F-statistic from first stage is equal to 23.2 but systematically less than 10 if we restrict the sample to the period 2004-12. We therefore recommend not using this instrument to derive the main results in this empirical application. It shall be clear that we use it only for illustrative purposes. The estimated PPWG is about 75%, much different from that estimated in Table 3, a sign of small relevance of the instrument or heterogeneous treatment effect .

A second strategy is to use one instrument and identify a set of admissible differentials for the entire population rather than for compliers only. Under the conditions introduced by Bhattacharya et al. (2008, 2012), further generalized by Chen et al. (2017) and Depalo (2018), the PPWG is bounded by $E[Y|Z=1] - E[Y|Z=0] \leq PPWG \leq E[Y|D=1, Z=1] - E[Y|D=0, Z=0]$.[10]

If the width of the set is "large", constant treatment effect is restrictive.[11] In our analysis, using the sector of the father as instrument the set is [0.150, 0.294] . Alternatively, using the sector of the mother as instrument the set is [0.171, 0.359] . Because both sets identify the PPWG for the entire population, their similarity supports the exogeneity of both instruments.[12]

4 Conclusion

Functional form assumptions play a small role in the estimation of the PPWG, but parametric (normality) assumptions introduce a substantial bias. Using either semi or nonparametric estimators, the gap is robustly estimated at about 20%, compared to 30% imposing normality.

[10]The conditions are those of the IV for Local Average Treatment Effect (Imbens and Angrist, 1994), as well as bounded outcome, rank similarity, and first order stochastic dominance.

[11]The exact meaning of "large" is an individual choice of the researcher that should trade-off the benefits from a simpler communication and the costs from unduly restrictive assumptions.

[12]The relevance of the instrument is less important for partial than for point identification, because the set is valid for the entire population rather than for compliers only; see Chen et al., 2017 for a thorough discussion.

5 Compliance with Ethical Standards

yes

References

Arabmazar, A. and P. Schmidt (1982). An investigation of the robustness of the tobit estimator to non-normality. *Econometrica 50*(4), 1055–1063.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008). Treatment Effect Bounds under Monotonicity Assumptions: An Application to Swan-Ganz Catheterization. *American Economic Review 98*(2), 351–56.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2012). Treatment effect bounds: An application to SwanGanz catheterization. *Journal of Econometrics 168*(2), 223–243.

Bound, J., D. Jaeger, and R. Baker (1995). Problems with instrumental variable estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*, 443–450.

Brunello, G. and C. Dustmann (1997). Le retribuzioni nei settori pubblico e privato in Italia e in Germania: un paragone basato su dati microeconomici. In C. Dell'Aringa (Ed.), *Rapporto ARAN sulle retribuzioni*, ARAN. Franco Angeli.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3 of *Handbook of Labor Economics*, Chapter 30, pp. 1801–1863. Elsevier.

Chen, X., C. A. Flores, and A. Flores-Lagunes (2017). Going beyond LATE: Bounding Average Treatment Effects of Job Corps Training. *Journal of Human Resources (forthcoming).*

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies 70*(1), 33–58.

Depalo, D. (2018). Identification Issues in the Public/Private Wage Gap with an Application to Italy. *Journal of Applied Econometrics 33*(3), 435–456.

Depalo, D. and R. Giordano (2011). The public-private pay gap: a robust quantile approach. *Giornale degli Economisti 70*(1), 25–64.

Dustmann, C. and A. van Soest (1998). Public and private sector wages of male workers in Germany. *European Economic Review 42*(8), 1417–1441.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica, Econometric Society 47*(1), 153–61.

Heckman, J. J. and E. J. Vytlacil (2007, January). Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6 of *Handbook of Econometrics*, Chapter 70. Elsevier.

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), 467–75.

Imbens, G. W. and D. B. Rubin (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies 64*(4), 555–574.

Lausěv, J. (2014). What has 20 years of publicprivate pay gap literature told us? eastern european transitioning vs. developed economies. *Journal of Economic Surveys 28*(3), 516–550.

Lucifora, C. and D. Meurs (2006). The public sector pay gap in france, great britain and italy. *Review of Income and Wealth 52*(1), 43–59.

Maddala, G. S. (1983). *Limited-Dependent and Qualitative Variables in Econometrics:*. Cambridge: Cambridge University Press.

Mourifié, I. and Y. Wan (2017). Testing local average treatment effect assumptions. *Review of Economics and Statistics 99*(2), 305–313.

Newey, W. K. (1999). Consistency of two-step sample selection estimators despite misspecification of distribution. *Economics Letters 63*(2), 129 – 132.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *Econometrics Journal 12*, S217–S229.

Nijman, T. and M. Verbeek (1992). Nonresponse in panel data: The impact on estimates of a life cycle consumption function. *Journal of Applied Econometrics 7*(3), 243–257.

Puhani, P. (2000). The heckman correction for sample selection and its critique. *Journal of Economic Surveys 14*(1), 53–68.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources 33*(1), 127–169.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), pp. 331–341.

Table 1: Descriptive Statistics

Variabile	Private Mean	Private S.D.	Private Median	Public Mean	Public S.D.	Public Median
Wage	1.889	0.582	1.966	2.107	0.594	2.161
Age	39.580	11.143	39.000	45.835	9.704	47.000
Low ed.	0.575	0.494	1.000	0.324	0.468	0.000
Middle ed.	0.350	0.477	0.000	0.414	0.493	0.000
High ed.	0.075	0.264	0.000	0.262	0.440	0.000
Blue collar	0.630	0.483	1.000	0.145	0.352	0.000
Manager	0.085	0.279	0.000	0.182	0.386	0.000
Partime	0.041	0.198	0.000	0.032	0.177	0.000
Father pub.	0.099	0.298	0.000	0.248	0.432	0.000
Mother pub.	0.054	0.225	0.000	0.087	0.282	0.000

Notes: sample size $N = 20,456$.

Table 2: Test of Instrument Validity - Father's Sector

Year	Education								
	Low			Middle			High		
	90	95	99	90	95	99	90	95	99
1998	NR	NR	NR	NR	NR	NR	NR	NR	NR
2000	NR	NR	NR	NR	NR	NR	NR	NR	NR
2002	NR	NR	NR	NR	NR	NR	NR	NR	NR
2004	NR	NR	NR	NR	NR	NR	NR	NR	NR
2006	NR	NR	NR	NR	NR	NR	NR	NR	NR
2008	NR	NR	NR	NR	NR	NR	NR	NR	NR
2010	NR	NR	NR	NR	NR	NR	NR	NR	NR
2012	R	R	R	NR	NR	NR	NR	NR	NR

Notes: the table reports the test of instrument validity as Mourifié and Wan (2017) at various confidence levels for each wave. 'R' stands for rejection of the null hypothesis; 'NR' for NON rejection.

Table 3: PPWG Estimates

Method	Estimates				
			Standard		
	OLS	IV			
PPWG	0.057 ***	0.284 ***			
	0.007	0.062			
			Parametric		
	Two-step	ML			
Heckman (1979)	0.319 ***	0.161 ***			
	0.032	0.021			
			Semiparametric		
Polynomial order	1	2	3	4	5
Newey (2009)	0.204 ***	0.203 ***	0.202 ***	0.202 ***	0.202 ***
	0.009	0.008	0.009	0.008	0.009
			Nonparametric		
Polynomial order	1	2	3	4	5
Das et al. (2003)	0.203 ***	0.202 ***	0.202 ***	0.202 ***	0.202 ***
	0.009	0.009	0.008	0.009	0.009

Notes: standard errors in parentheses; *** denotes significance at the 99% confidence level; sample size $N = 20,456$. The sample for the semiparametric and nonparametric approaches is trimmed 1% on both sides.

Table 4: Estimated Coefficients

	OLS	IV	Heckman (1979)		Newey (2009)					Das et al. (2003)				
			Two-step	ML	1	2	3	4	5	1	2	3	4	5
Age	0.009 ***	0.007 ***	0.007 ***	0.008 ***	0.008 ***	0.008 ***	0.008 ***	0.007 ***	0.008 ***	0.014 ***	0.013 ***	0.012 ***	0.012 ***	0.012 ***
	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.001	0.001
Age sq.	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***	-0.000 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Married	0.098 ***	0.093 ***	0.092 ***	0.096 ***	0.088 ***	0.086 ***	0.085 ***	0.084 ***	0.084 ***	0.112 ***	0.102 ***	0.098 ***	0.098 ***	0.100 ***
	0.007	0.007	0.007	0.007	0.006	0.006	0.006	0.006	0.006	0.026	0.026	0.026	0.026	0.026
Low ed.	-0.243 ***	-0.205 ***	-0.199 ***	-0.226 ***	-0.223 ***	-0.196 ***	-0.184 ***	-0.175 ***	-0.179 ***	-0.229 ***	-0.195 ***	-0.181 ***	-0.182 ***	-0.187 ***
	0.010	0.015	0.012	0.010	0.008	0.010	0.011	0.011	0.012	0.034	0.034	0.035	0.035	0.035
Middle ed.	-0.162 ***	-0.120 ***	-0.113 ***	-0.142 ***	-0.146 ***	-0.122 ***	-0.109 ***	-0.099 ***	-0.103 ***	-0.109 ***	-0.074 **	-0.060 *	-0.060 *	-0.065 **
	0.009	0.015	0.011	0.010	0.008	0.009	0.011	0.011	0.012	0.030	0.031	0.032	0.032	0.032
Blue collar	-0.133 ***	-0.064 ***	-0.053 ***	-0.101 ***	-0.132 ***	-0.164 ***	-0.137 ***	-0.124 ***	-0.126 ***	-0.312 ***	-0.289 ***	-0.251 ***	-0.247 ***	-0.238 ***
	0.007	0.020	0.012	0.009	0.007	0.011	0.017	0.017	0.017	0.039	0.040	0.044	0.045	0.045
Manager	0.254 ***	0.276 ***	0.279 ***	0.264 ***	0.226 ***	0.232 ***	0.239 ***	0.244 ***	0.243 ***	0.244 ***	0.262 ***	0.271 ***	0.271 ***	0.268 ***
	0.009	0.011	0.010	0.009	0.008	0.008	0.009	0.009	0.009	0.032	0.033	0.033	0.033	0.033
Part-time	-0.039 ***	-0.042 ***	-0.043 ***	-0.040 ***	-0.024 **	-0.024 **	-0.024 **	-0.025 **	-0.025 **	-0.070	-0.072	-0.078	-0.078	-0.078
	0.013	0.014	0.014	0.013	0.011	0.011	0.011	0.011	0.011	0.060	0.059	0.060	0.060	0.060

Notes: standard errors in parentheses; *** denotes significance at the 99% confidence level; sample size $N = 20,456$. The sample for the semiparametric and nonparametric approaches is trimmed 1% on both sides.

Table 5: Test of Instrument Validity - Mother's Sector

Year	Education								
	Low			Middle			High		
	90	95	99	90	95	99	90	95	99
1998	R	NR	NR	NR	NR	NR	NR	NR	NR
2000	NR	NR	NR	R	NR	NR	NR	NR	NR
2002	R	R	R	R	R	R	NR	NR	NR
2004	R	R	NR	R	R	NR	NR	NR	NR
2006	R	R	R	R	R	R	NR	NR	NR
2008	R	R	R	NR	NR	NR	R	R	R
2010	R	R	NR	NR	NR	NR	NR	NR	NR
2012	NR	NR	NR	R	R	NR	NR	NR	NR

Notes: the table reports the test of instrument validity as Mourifié and Wan (2017) at various confidence levels for each wave. 'R' stands for rejection of the null hypothesis; 'NR' for NON rejection.

A Asymptotic Bias from Parametric Assumptions

Let the true model be given by equations 1-3, and the true distribution of ε_{ji} be unknown for $j = \{0,1\}$. Define $\theta_j \equiv (\beta_j, \xi_j)$ and $\omega_{ji} \equiv (x_i', \lambda_j(z_i))'$, where $\lambda_j(z_i)$ is the correction term ($\lambda_0(z_i) = \frac{\phi(z_i'\gamma)}{1-\Phi(z_i'\gamma)}$ and $\lambda_1(z_i) = \frac{\phi(z_i'\gamma)}{\Phi(z_i'\gamma)}$). It is possible to rewrite the outcome equation as $y_{ji} = \omega_{ji}'\theta + u_{ji}$, where $u_{ji} = \varepsilon_{ji} - \xi_j \lambda_j(z_i)$, which is a zero-mean error term. Hence, it can be interpreted in terms of the Heckman two-step estimator. The probability limit of this estimator is given by:

$$\hat{\theta}_j^{H2S} \equiv \left(\frac{1}{N} \sum_{i=1}^{N} \omega_{ji}\omega_{ji}' \right)^{-1} \frac{1}{N} \sum_{i=1}^{N} \omega_{ji} y_{ji} \xrightarrow{P} \theta_j + \mathbb{E}\left[\omega_{ji}\omega_{ji}'\right]^{-1} \mathbb{E}\left[\omega_{ji} u_{ji}\right]$$

Even though the mean regression model is correctly specified, β_j may suffer from an asymptotic bias equal to:

$$ABias\left(\hat{\theta}_j^{H2S}\right) = \mathbb{E}\left[\omega_{ji}\omega_{ji}'\right]^{-1} \mathbb{E}\left[\omega_{ji}\mathbb{E}\left(\varepsilon_{ji}|d_i = j, z_i\right) - \omega_{ji}\xi_j\lambda_j(z_i)\right]$$

Hence, the severity of the bias is proportional to $\mathbb{E}\left[\omega_{ji}\mathbb{E}\left(\varepsilon_{ji}|d_i = j, z_i\right) - \omega_{ji}\xi_j\lambda_j(z_i)\right]$, which equals 0 only under normality (if an exogenous and relevant instrument exists).

B Additional table (available on the website)

Table 6: Stability of the Estimates to the Inclusion of Covariates

Method	1	2	3	4
	Age Edu. Year	1 + Respons.	2 + Part time	All
Public	0.325 ***	0.277 ***	0.278 ***	0.284 ***
	0.052	0.058	0.058	0.062

Notes: standard errors in parentheses; *** denotes significance at the 99% confidence level.