Identification issues in the public/private wage gap, with an application to Italy

Domenico Depalo

Economics and Statistics Department, Banca d’Italia, Rome, Italy

Correspondence
Domenico Depalo, Economics and Statistics Department, Banca d’Italia, Via Nazionale 91, 00184 Rome, Italy.
Email: domenico.depalo@bancaditalia.it

Summary
This paper reviews some of the standard assumptions that are imposed in order to estimate the average public/private wage gap and that are mainly related to the possible selection of the sector. There are two contributions to the existing public/private wage gap literature. One is a better understanding of the identified parameters: standard estimators identify a local effect (LATE), which in general cannot be generalized to the entire population, as instead is almost always done. The other is the partial identification of the population average treatment effect, with an instrumental variable. To the best of my knowledge, this is the first paper in this literature that employs bounds. The technique is applied to male workers in Italy. For compliers, LATE estimates a wage advantage from working in the public sector greater than 30%. This return is within the narrowest bounds on the population average treatment effect that are consistent even with a much smaller gap (about 15% or more).

1 | INTRODUCTION

The public/private wage gap enjoyed a renewed interest during the recent economic recession, when some countries cut public sector wages (either in nominal or real terms) in an attempt to restore their fiscal positions. Understanding (1) whether a wage gap in favor of public sector workers still exists with respect to the private sector counterparts, (2) under what circumstances and (3) how large it is, are key questions for policymakers and researchers.

This paper tackles these issues by proposing two contributions to the existing public/private wage gap literature: The first is a better understanding of the point identification based on standard techniques; the second is a different approach, never investigated before in this literature, that exploits mild nonparametric assumptions, in addition to an instrumental variable (IV), to identify the population average wage gap. To this aim, I estimate bounds instead of single points.

The incredibly large existing body of literature on the public/private wage gap typically employs standard techniques without questioning the underlying hypotheses. Very few papers discuss the plausibility of necessary assumptions and even fewer cast doubt on (at least some of the) conclusions. Almost always, the choice is between sorting based on observable or unobservable (to the researcher) characteristics, with a preference for the latter over the former. Seminal papers imposed the assumption of independence between sector of employment and outcome (selection at random or ignorable; Smith, 1976). However, whether the employee works in the private or in the public sector can be the result of an individual utility maximization; thus the independence assumption might be unduly restrictive and accordingly has been removed in a number of studies (Dustmann & van Soest, 1998; is probably the most complete treatment of this issue). When the sorting is based on unobservable characteristics and an IV approach is used, in general the wage gap can be identified only for the subpopulation made up of individuals whose sector of employment is affected by the instrument (or compliers in the terminology of Angrist, Imbens, & Rubin, 1996). To the best of my knowledge, this has never been emphasized before in this literature.
By estimating a range of values (“bounds”), it is possible to allow for sorting based on unobservable characteristics and still (partially) identify the average treatment effect for the population, using an IV. The more homogeneous the parameter within the population, the smaller the range.

Bounds are becoming increasingly popular for empirical studies. The methodological research initiated by Manski (1990) moves away from the conventional focus on models that allow point identification, in favor of partial identification (Manski & Nagin, 1998). One of its goals is to “explore the estimates delivered by different sets of assumptions without the need to make enough … assumptions to achieve point identification” (Ho & Rosen, 2015, p. 3). Bounds provide a clear connection between the structure imposed on the data through assumptions, credibility of results that follow and precision of the conclusions that may be drawn (whereby the smaller the set of estimated effects, the higher the precision). However, it is only recently that the advantage of exploring a variety of assumptions is properly weighted against the disadvantage of having a set of admissible values for the parameter of interest instead of a single point.¹ To the best of my knowledge, this paper is the first application of bounds in the public/private wage gap literature.

In the available analyses, based on point identification, in almost all developed countries a positive average wage gap is estimated in favor of the public sector workers, although with notable cross-country heterogeneity (Giordano et al., 2014). When possible sorting is considered, the pay gap usually increases by as much as 30 percentage points for men. No further qualification is provided beyond this number.

In this paper I focus on Italy, which is a particularly interesting case to analyze because in 2010 a wage freeze for the public sector was introduced for period 2011–2016. Following a large part of the literature, I consider only men.² In line with the existing results, by imposing random selection I estimate a wage advantage for public sector workers that is as large as 5%. If some sorting mechanism is in operation, the pay gap estimated by IV is much larger, above 30%: As this paper emphasizes, this is the effect for compliers only. Using the narrowest bounds on the population average treatment effect of Bhattacharya, Shaikh, and Vytlacil (2012), the data are coherent also with a smaller pay gap (about 15% or more), which is still larger than that estimated under ignorable selection. Although wide bounds may be unpleasant for the policymaker, who typically prefers a single number summarizing everything, they reflect the importance of heterogeneity in the working population (Horowitz & Manski, 2000). Not only are upper and lower bounds informative, but also their “distance” is informative. In this respect, a relevant feature of bounds is the possibility to precisely distinguish the role of assumptions (or “assumption effect”) from that of the population to which they apply (or “population effect”).

The paper is organized as follows. In Section 2, I briefly present a motivating example that shows how critical the random sorting assumption may be. I then analyze the techniques that will be used in the empirical application (Section 3), focusing on their advantages and disadvantages. After a brief description of the pay gap in Italy from the National Accounts data (Section 4), I apply the methods in Section 5. In Section 6, I offer some conclusions.

2 | A MOTIVATING EXAMPLE

With only a few exceptions (e.g., Forni & Giordano, 2003; Holmlund, 1993), the existing literature in this field is mainly empirical in its nature. Even though this paper is not an exception, a motivating example may be of help to introduce the approaches adopted in the following sections and to interpret the results.

A worker may work in the public or in the private sector. The status is mutually exclusive. Early studies on the pay gap made the implicit assumption that workers are indifferent between the two sectors, so that the choice is made at random. In a more realistic setting, coherent with the Roy (1951) model, workers decide upon the employment sector that maximizes their utility. Heckman and Honoré (1990) provide further generalizations of the original model. Most importantly, this breaks down the hypothesis of independence between sector and wage and requires an appropriate empirical approach that is analyzed in Section 3.

Maximization may depend on complex mechanisms, including the nonmonetary private benefit $B$ that is associated with the job the worker does (for surveys see Gregory & Borland, 1999; Lausèv, 2014). The benefit is unobservable to the researcher, but the existing literature thinks of it as depending on a broad definition of motivation, risk aversion and similar attitudes. For example, if $B$ depends on a dichotomous indicator for motivated versus nonmotivated workers to

¹Although the wording reflects the conventional wisdom, I do not believe that a range is a “disadvantage” per se.
²Indeed, “for females the participation decision should be taken into account, and this requires a different model” (Dustmann & van Soest, 1998, p. 1419). For Italy, this is an even more concerning argument, because, according to ILO statistics (http://www.ilo.org/global/statistics-and-databases/lang--en/index.htm), nonactive women are about 50%, one of the highest shares around Europe. For comparison, nonactive men are about 20%, similar to other European countries.
work in the public sector, then the probability of working in the public sector is higher for the former than for the latter group because the benefit is higher. Associated with benefits there are opportunity costs: Continuing with the example of motivation, if the ability of the worker is high or if he applies for jobs with limited responsibilities, the opportunity cost to join the public sector (which is based on a competition to be admitted) would be lower and the individual utility will be higher. The interpretation in terms of cost–benefit analysis will be of much help to interpret the results and will be further elaborated upon in Section 5.

3 | EMPIRICAL STRATEGY

In this section, I describe the techniques used in the empirical application. To reduce notation, but without loss of generality, I do not explicitly condition on observable characteristics (X); however, everything should be conditioned on them. In contrast to the existing literature on the public/private wage gap, I find it convenient to consider the sector of employment as a treatment. Borrowing from that literature, a better understanding of what quantities are identified and under which conditions is immediate. This greatly improves over existing results. Define Y as the wage and Yd the potential wage a worker earns in sector d ∈ D, an indicator equal to 1 for the public sector and 0 for the private sector. Potential wages are pairs of outcomes defined for the same worker, given different exposure to treatment. The ultimate goal of the analysis is the evaluation of the average treatment effect (ATE) for the population of interest:

\[ \Delta_y = E[Y_1] - E[Y_0] \]
\[ = \{E[Y_1|D = 1] P(D = 1) + E[Y_1|D = 0] P(D = 0)\} \]
\[ - \{E[Y_0|D = 1] P(D = 1) + E[Y_0|D = 0] P(D = 0)\}, \]

that is, the expected gain from working in the public sector for a randomly chosen worker (Heckman, Tobias, & Vytlacil, 2003). If I had the opportunity to observe the outcome under both treatment states for the same individual, the estimation would be straightforward. This is the approach in Disney and Gosling (2008). However, in general, for each worker I can observe either Y0 or Y1. In these situations, the key issue is recovering the wage in the unobservable status. As emphasized Imbens and Wooldridge (2009), the potential outcomes framework clarifies where the uncertainty in the estimators comes from, a virtue that I exploit below.

3.1 | Standard approach

Under ignorable selection (i.e., \( D \perp (Y_0, Y_1) \)), sector sorting is not an issue; therefore \( E[Y_d|D = 0] = E[Y_d|D = 1] \) and substituting in Equation 1 it follows that \( \Delta_y = E[Y_1|D = 1] - E[Y_0|D = 0] \). This restriction implies that an ordinary least square (OLS) of wage on treatment indicator is consistent for the average wage gap in the population (ATE), provided the wage setting in the two sectors is equal up to a location shift (Imbens & Wooldridge, 2009).

If workers sort on the basis of their unobservable preferences, such that the selection is nonignorable (Little, 1995), the OLS is an inconsistent estimator for the wage differential. Solving this drawback has been central in the existing literature. Standard approaches involve an IV estimator, based on an instrument \( z \in Z \) taking values 0 or 1, that affects the decision of the sector, but not the wage. To formalize the role of the instrument, define potential wages as \( Y(z, d) \) and potential treatment status as \( D \) when \( Z = 1 \) and \( D_0 \) when \( Z = 0 \), respectively. The decision mechanism is a flexible threshold-crossing model:

\[ D = 1(\sigma(Z, v) > 0), \]

where \( v \) is a disturbance and \( 1(A > 0) \) is an indicator function taking value 1 if \( A > 0 \).

Within this framework, an IV identifies a local average treatment effect (LATE) as proposed by Imbens and Angrist (1994) and Angrist et al. (1996). If the following assumptions hold: (1) the potential wages for each worker are unrelated to the treatment status of other workers (stable unit treatment value assumption); (2) the instrument is randomly assigned; (3) exclusion restriction (i.e., \( Y_d \equiv Y(0, d) = Y(1, d) \)); (4) nonzero average causal effect of \( Z \) on \( D \) (i.e., \( E[D_1 - D_0] \neq 0 \)); (5) monotonicity (i.e., \( D_1 \geq D_0 \) for all workers, such that an increase in the level of the instrument does not decrease the level of the treatment; or vice versa); then the IV-LATE is:

\[ \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} = E[Y_1 - Y_0|D_1 > D_0]. \]
This is the treatment effect for compliers, that is, workers who are induced to work in the public sector by a change in the instrument. This quantity in general does not identify the average treatment effect for the entire population, because different instruments induce the change in the treatment for different subpopulations. As a consequence, using different instruments in general leads to different estimations of marginal return, which may be above or below the OLS estimates, each valid for different subpopulations. This well-known result has not been emphasized before in the public/private wage gap literature, although it is key to correctly interpret the estimates. This is one contribution of the paper. The relevance of the IV-LATE parameter has been discussed at length in the policy evaluation literature (Deaton, 2010; Heckman & Urzúa, 2010; Imbens, 2010), but ultimately it depends on the empirical context (Huber, Laffers, & Mellace, 2017). Yet, beyond the circumstance that the entire population is made of compliers, there are important special cases where the IV-LATE has a more general empirical content: One is when the pay gap is identical across all the subpopulations (or “common coefficient model”); one is when the wage gap varies across subpopulations but workers do not select the sector on the basis of idiosyncratic component of their return, in which case the IV identifies the mean effect of treatment on the treated or on randomly selected workers (Heckman, 1997). Implicitly, the existing literature always imposes one of these hypotheses; which of them is actually invoked cannot be said, because explicitly it is never stated.

Exploiting the local identification power of IV-LATE, Ichino and Winter-Ebmer (1999) propose an intriguing solution to estimate the return on education in the population of Germany. The idea underlying their approach is that if one could find instruments capable of identifying the highest and the lowest returns the corresponding estimates would allow us to bracket the range of variation of the public/private wage gap in the population. This represents a simple and important departure from point estimation.

Finally, violation of the exclusion restriction assumption induces a bias in unknown direction; also, the violation of monotonicity leads to a bias in unknown direction due to the existence of defiers, that is, individuals who do the opposite with respect to what the instrument assignment would imply (Angrist et al. 1996). Despite their critical role, the validity of these assumptions has been taken for granted in the literature and never tested. In the empirical application, I fill this important gap.

### 3.2 A different approach: Bounds

In order to identify the population, rather than the local, average treatment effect even with an IV, other solutions must be investigated. If only mild assumptions—such as those reviewed in this subsection—are imposed, this usually comes at the cost of losing point identification.

The overall difficulty of the analysis is that I do not observe \( Y_1 \) for private sector workers and \( Y_0 \) for public sector workers. Using basic statistical tools, Manski (1990) introduced a perspective identifying a set of admissible marginal effects that depend on the underlying assumptions. Stronger assumptions on the unobserved components narrow the bounds. A virtue of the approach is that hypotheses are clearly stated, so that one may check (i) whether they are credible or not and (ii) whether some more structure may be imposed or not.

In the simplest analysis, for the partial identification of the population average treatment effect it is only required that the support \( Y \in [k_0, k_1] \) (hence they are “bounded outcome” bounds). The lower (upper) bound of the wage level can be obtained after the substitution of \( k_0 (k_1) \) in the unobservable status: The only feature that I know is that \( Y \) is at least equal to \( k_0 \) (at most equal to \( k_1 \)), but not exactly how much it would be if I could observe it; by the same token, the lower bound on the population average treatment effect is the difference between the lower bound on \( E[Y_1] \) and the upper bound on \( E[Y_0] \) and vice versa for the upper bound on the population average treatment effect. This explanation highlights a deep difference between bounds and point estimators equipped with confidence intervals: With the latter, I know that the likelihood of the parameter being near the center of the confidence interval is a lot larger than it being near the boundaries (loosely speaking); with the former, I do not know a priori where the true parameter may be.3 “Bounded outcome” bounds are not presented in the empirical analysis, but it may be instructive to look at their analytical form to have a more concrete idea of the above explanation about how the method works. From Equation 1, it follows that the bounds on the population average treatment effect are

\[
\text{Bounded outcome bounds:}
\]

\[
\begin{align*}
\text{lower: } & [E[Y|D = 1] P(D = 1) + k_0 P(D = 0)] - [k_1 P(D = 1) + E[Y|D = 0] P(D = 0)] \\
\text{upper: } & [E[Y|D = 1] P(D = 1) + k_1 P(D = 0)] - [k_0 P(D = 1) + E[Y|D = 0] P(D = 0)].
\end{align*}
\]

---

3This explanation has been suggested by a referee.
The width of these bounds, obtained as the difference between the upper and the lower bound, is equal to \((k_1 - k_0)\); that is, the larger the admissible values, the larger the width. This is the first, not really satisfactory, indicator of heterogeneity in the population.

The existing literature on the public/private wage gap may suggest assumptions to narrow the bounds (Gregory & Borland, 1999; Lausėv, 2014). For example, a situation where each person’s wage function is higher in the public than in the private sector is consistent with the monotone treatment response (MTR; Manski, 1997; Manski & Pepper, 2000), which implies that \(Y_1 \geq Y_0\) for all workers (Manski, 1990). Under this MTR, the bounds on the population average treatment effect are

\[ \text{MTR bounds:} \]
\[ \begin{align*}
\text{lower:} & \quad 0 \\
\text{upper:} & \quad E[Y|D = 1] - [k_0 \ P(D = 1) + E[Y|D = 0] \ P(D = 0)].
\end{align*} \tag{5} \]

Since the lower bound implied by the MTR assumption is equal to 0 (Manski, 1997, Proposition M2, p. 1320), it is only the upper bound that deserves further investigation. The upper bound from MTR is smaller than the bounded outcome upper bound by an amount proportional to the distance between the largest admissible value of the wage and the average wage for the treated (i.e. \(k_1 - E[Y_1|D = 1]\)). This reduction may be understood as the gain from this specific assumption.

So far, I never used an instrument. If there exists one, such that it is assumed that \(Y_0\) and \(Y_1\) are independent of \(Z\), then—using also “bounded outcome” and Equation 2—Manski (1990) and Heckman and Vytlacil (2001) show that the bounds on the population average treatment effect are

\[ \text{Independence bounds:} \]
\[ \begin{align*}
\text{lower:} & \quad [E[Y|D = 1, Z = 1] \ P(D = 1|Z = 1) + k_0 \ P(D = 0|Z = 1)] \\
& \quad - [k_1 \ P(D = 1|Z = 0) + E[Y|D = 0, Z = 0] \ P(D = 0|Z = 0)] \\
\text{upper:} & \quad [E[Y|D = 1, Z = 1] \ P(D = 1|Z = 1) + k_1 \ P(D = 0|Z = 1)] \\
& \quad - [k_0 \ P(D = 1|Z = 0) + E[Y|D = 0, Z = 0] \ P(D = 0|Z = 0)].
\end{align*} \tag{6} \]

With an IV, these bounds employing the assumption that the outcome is bounded are the most basic to deal with the fact that \(Y_1\) is never observed for individuals who never work in the public sector regardless of the value of the IV (called “never takers”) and \(Y_0\) is never observed for individuals who always work in the public sector regardless of the value of the IV (called “always takers”). See Chen, Flores, and Flores-Lagunes (2017) for a formal statement of the problem, with close connection to this paper. Independence bounds can be narrowed if further assumptions are imposed: those explored below are nonparametric in nature, but stronger that those in Manski (1990). Shaikh and Vytlacil (2011), henceforth SV, impose (1) that \(E[s(1, v)] \geq E[s(0, v)]\) or vice versa; that is, the treatment is monotone in the instrument although in an unknown direction—consistent with the idea that workers more motivated to work in the public sector are also more likely to work in this sector than in the private—and (2) the rank similarity (Chernozhukov & Hansen, 2005), that is, a subpopulation defined by given characteristics \((x = X, z = Z)\), shows the same distribution of ranks across treatment states (Frandsen & Lefgren, 2017), or formally \(\epsilon_d|v \sim \epsilon,\vbar\) with \(\epsilon_d\) an unobserved disturbance of \(Y_d\). Under these circumstances, if \(E[Y|Z = 1] > E[Y|Z = 0]\) (as in the empirical application of Section 5), the bounds on the population average treatment effect are

\[ \text{SV bounds:} \]
\[ \begin{align*}
\text{lower:} & \quad E[Y|Z = 1] - E[Y|Z = 0] \\
\text{upper:} & \quad E[Y|D = 1, Z = 1] \ P(D = 1|Z = 1) + k_1 \ P(D = 0|Z = 1) \\
& \quad - E[Y|D = 0, Z = 0] \ P(D = 0|Z = 0) - k_0 \ P(D = 1|Z = 0). \tag{7}
\end{align*} \]

SV bounds (i) are sign defined because, if the lower bound is positive, then the upper bound will be “more” positive (or vice versa, if the upper bound is negative the lower bound will be “more” negative). Compared to the MTR assumption

Manski and Pepper (2000) introduce also the monotone treatment selection (MTS) assumption. It implies that those who work in the public sector have a weakly higher mean wage function than those who work in the private sector. MTR and MTS can be exploited together (the empirical application is run imposing also this assumption; the interested reader will find it on my website).

More precisely, Shaikh and Vytlacil (2011) impose the equality of error distributions in the wage equation between the two sectors. This restriction has been weakened by the rank similarity in Bhattacharya et al. (2012): following the latter paper, I still refer to these bounds as SV to avoid confusion with the BSV bounds presented below. Also, the original bounds in Shaikh and Vytlacil (2011) and Bhattacharya et al. (2012) apply to the binary outcomes, whereas the bounds exploiting the mean independence of \(Z\) (Manski, 1990) or MTR (Manski & Pepper, 2000) apply to generic outcomes; therefore, when necessary, I derived the bounds for a generic outcome (see also Chen et al., 2017).
TABLE 1 Hypotheses for each bound. Case of $E[y|Z = 1] > E[y|Z = 0]$

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Bound outcome</th>
<th>MTR</th>
<th>Indep.</th>
<th>SV</th>
<th>BSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>$y \in [k_0, k_1]$</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>$y_1 \geq y_0, \forall i$</td>
<td></td>
<td></td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank similarity</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold crossing</td>
<td>V</td>
<td>V</td>
<td>V</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PQD</td>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td></td>
</tr>
</tbody>
</table>

of Manski and Pepper (2000), where one knows a priori that $Y_1 \geq Y_0$ for all workers (or vice versa), SV bounds (ii) allow the effect to be positive for some workers and negative for others, and (iii) identify the sign of the population average treatment effect from the data; conversely, MTR bounds do not use an IV and do not impose the threshold crossing model. Bhattacharya et al. (2012) (WP version) show that, (iv) under the economic framework outlined in Section 3.1, the bounds of the analysis of Manski and Pepper (2000) with MTR with $Y_1 \geq Y_0$ (or vice versa) and the instrumental variable assumption of Manski (1990), wages are independent of $Z$, simplify, and coincide with the analysis in Shaikh and Vytlacil (2011), when $E[Y|Z = 1] > E[Y|Z = 0]$ (or vice versa). Finally, (v) SV bounds are narrower than independence bounds of Equation 6.

Furthermore, in the existing literature a wage gap in favor of the public sector is estimated at all quantiles of the wage distribution: for Italy, see Lucifora and Meurs (2006) and Depalo and Giordano (2011); for several EU countries, see Giordano et al. (2014). This implies a form of stochastic dominance (the positive quadrant dependence, PQD); that is, two variables $\epsilon$ and $v$ are more likely to be large together or to be small together compared to $\epsilon'$ and $v'$, where $\epsilon \sim \epsilon'$ and $v \sim v'$, and $\epsilon'$ and $v'$ are independent of each other (Joe, 1997) or, formally, $P(\epsilon \leq t_0|v \leq t_1) \geq P(\epsilon \leq t_0), \forall(t_0, t_1)$. Imposing this structure, Bhattacharya et al. (2012) show that if $E[Y|Z = 1] > E[Y|Z = 0]$ the bounds on the population average treatment effect may further shrink to

**BSV bounds**:

\[
\text{lower: } E[Y|D = 1, Z = 1] - E[Y|D = 0, Z = 0],
\]

which are narrower than those in Shaikh and Vytlacil (2011) thanks to a smaller upper bound. The width of these bounds may also be negative, which would be a strong evidence against underlying hypotheses.

Before proceeding, it may be useful to clarify that SV and BSV bounds, on which the main results of Section 5 are based, and IV-LATE, remove the hypothesis of selection at random and rely on a suitable IV (in particular, with respect to exogeneity and monotonicity assumptions); but SV and BSV bounds identify the population average treatment effect, while IV-LATE, in general, point identifies the effect for compliers only. Therefore, it may also happen that IV-LATE is consistent, but outside the SV or BSV bounds on the population average treatment effect.

Table 1 summarizes the required hypotheses to identify each bound.

Early treatment of bounds is in Manski (1990), whereas a general approach can be found in Manski (2003). Extensions of bounds used in this paper are in Chen et al. (2017). Some fields where partial identification has been used are labor market (e.g., Chen et al., 2017; Lee, 2009), health (e.g., Bhattacharya, Shaikh, & Vytlacil, 2008), (2012), for the effect of catheterization; Gundersen, Kreider, & Pepper, (2012), or Gundersen & Kreider, (2009), to evaluate children’s health), insurance (e.g., Kreider & Hill, (2009), schooling (e.g., Blanco, Flores, & Flores-Lagunes, 2013; Huber et al., 2017), wage inequality (e.g., Blundell, Gosling, Ichimura, & Meghir, 2007), crime (e.g., Manski & Pepper, 2013), and domestic violence (e.g., Siddique, 2013). This is the first paper using bounds in the public/private wage gap literature.

4 | THE PAY GAP IN THE AGGREGATE NATIONAL ACCOUNTS DATA

To better appreciate the policy relevance of this paper, it is worth emphasizing that according to aggregate National Accounts data the payment conditions in the public sector are better than in the private sector. Depalo and Giordano (2011) document that the difference in pay between the two sectors in recent decades has always been sizable. The gap
was about 20% in 1980 and reached almost 40% in 1990, following a series of particularly favorable wage renewal contracts in the public sector; it decreased to about 20% in 1995, reflecting the overall fiscal consolidation effort required under the Maastricht Treaty to join the European Monetary Union; the differential started increasing again at the beginning of the last decade until 2010 (about 40%). The Budgetary Law issued at the aftermath of the crisis in 2010 (Law 78/2010 and the late modifications) introduced several measures to contain the public sector wage bill. Among them were: a wage freeze at the wage level of fiscal year 2010 for all the public sector employees that—due to the late law modifications—lasted for 5 years (from 2011 to 2016); the reduction of the highest wage levels and the block of contractual renewals (Banca d’Italia, 2011; http://www.funzionepubblica.gov.it/media/infografiche/riforma-della-pa/ 01-12-2016/riforma-pa-e-nuovo-contratto; a new agreement was reached at the end of 2016); severe rules for turnover of the workforce that affected, overall, the youngest cohorts (Banca d’Italia, 2016). As a consequence, the differential started decreasing, although only marginally, reaching about 35% in 2012 (Sestito, 2017). More details on the public sector in Italy and the related literature are in Appendix A1.

5 RESULTS

The techniques in Section 3 are applied to data obtained from the Survey on Household Income and Wealth (SHIW), conducted every 2 years by the Bank of Italy on a sample representative of the Italian population. This dataset has been largely employed to study the public/private pay gap in Italy. The reference period for the analysis is 2006–2012. Further details on the data and descriptive statistics are given in Appendix A2.

I begin the analysis with the classical approach under ignorability, which is quickly relaxed in order to allow for sorting of the sector (Section 5.1). Although there is nothing new in the latter approach, to the best of my knowledge this is the first attempt to formally support the sector of the father as instrument and to properly address a tight identification of the parameters in this field.

The novelty of this paper is in the abandoning of point identification in favor of bounds (Section 5.2), in order to identify the population average treatment effect even with an IV. The data are consistent with “some structure”; therefore, I nonetheless obtain economically relevant results.

5.1 Yet another estimate: Standard approaches

As in the existing literature, in the standard approaches the conditioning set includes a second-degree polynomial in age and a set of dummies for educational attainment (basic, low, and high level), marital status, job responsibilities (blue-collar workers and managerial position), and geographical areas (four distinct macroregions). When the waves are pooled, a set of time dummies is added.

5.1.1 OLS-ATE

Under random sector selection, in pooling all the waves, in the population there is an average 5% wage gap in favor of public sector workers (Table 2); by year, the range is between 5% and 8% (when significant, at standard confidence level). The gap is very erratic over time and a clear path cannot be identified: It is at about 7% in 2008 and 2012, and and it is a lower percentage in other waves. Despite the wage freeze in the public sector that was introduced by the Budgetary Law since 2011 (Section 4), between 2010 and 2012 the wage gap remained virtually constant, in the range of variation observed in previous years. The likely reasons, although they are not mutually exclusive, are that during these two years the wages in the private sector diminished further because of the dramatic economic downturn, or that in only one year the policy did not exert its effects completely.

5.1.2 IV-LATE

The example in Section 2 suggests that workers may decide upon the sector that maximizes their utility. The sorting mechanism is unknown to the researcher: Factors that influence the decision are thought to be motivation, ability, and risk aversion. SHIW allows us to explore the motivation channel, using the fathers’ sector of occupation (see Dustmann & van Soest, 1998; and, for Italy, Bardasi, 1996, and Brunello & Dustmann, 1997). The exposition of results from IV-LATE consists of two parts: First I discuss the validity of the instrument, then I present the wage gap.
I analyze the legitimacy of the instrument from an economic and a statistical perspective. There is a well-established tradition of using the fathers’ sector of occupation as instrument for public sector sorting, on the rationale that the sector of the father is uncorrelated with the individual’s wage level (i.e., exogenous), but it is relevant because it shapes the preferences of an individual for choosing to work in one sector or another (Depalo & Giordano, 2011), as the preference for the public sector exhibits an intertemporal persistence from fathers to children (Brunello & Dustmann, 1997; Dustmann & van Soest, 1998).

In order to check whether these economic arguments are coherent with statistical arguments, I test the instrument validity using recent approaches proposed by Kitagawa (2015) (see also Huber & Mellace, 2015) and (Mourifié & Wan 2016). Both are based on results in Heckman and Vytlacil (2005). It should be borne in mind that nonrejection of these null hypotheses does not confirm the IV validity, which thus is a refutable but nonverifiable hypothesis. For the case of a binary instrument Z, the assumptions tested are those related to exclusion restriction and monotonicity; the relevance of the instrument is not tested and relies on the first-stage statistics. Let the probability distributions be \( P(y, d) = \Pr(Y = y, D = d|Z = 1) \) and \( Q(y, d) = \Pr(Y = y, D = d|Z = 0) \): If the instrument is valid, \( P(y, 1) \) must nest \( Q(y, 1) \) for treatment outcomes and \( Q(y, 0) \) must nest \( P(y, 0) \) for control outcomes; otherwise, if density estimates intersect, at least one of the assumptions is refuted. Using the father’s sector of occupation as instrument, from Figure 1 the estimated densities overall exhibit a nesting relationship. Also, I implement the test proposed by Mourifié and Wan (2016). It tests the implications of the LATE assumptions about the exclusion restriction and monotonicity assumptions. These implications take the form of two inequalities: If either of the two inequalities is violated, the test rejects the joint assumptions required for the IV-LATE and the validity of the instrument is falsified. When the instrument is the father’s sector of occupation, at standard confidence levels the null hypothesis is not rejected (Table 3). Therefore, the validity of the sector of the father as instrument is not falsified. As a comparison run for illustrative purposes, using a different potential instrument, namely

### Table 2: Estimated wage gap: various strategies as declared at the head of each panel

<table>
<thead>
<tr>
<th>Gap</th>
<th>2006</th>
<th>2008</th>
<th>2010</th>
<th>2012</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS-ATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.024</td>
<td>0.070***</td>
<td>0.049***</td>
<td>0.079***</td>
<td>0.054***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.010; 0.057]</td>
<td>[0.038; 0.102]</td>
<td>[0.014; 0.083]</td>
<td>[0.043; 0.116]</td>
<td>[0.037; 0.071]</td>
</tr>
<tr>
<td>Obs.</td>
<td>3,200</td>
<td>3,181</td>
<td>2,962</td>
<td>2,893</td>
<td>12,236</td>
</tr>
<tr>
<td>IV-LATE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.144</td>
<td>0.467**</td>
<td>0.254*</td>
<td>0.435**</td>
<td>0.333***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.364; 0.652]</td>
<td>[0.031; 0.903]</td>
<td>[−0.032; 0.539]</td>
<td>[0.021; 0.849]</td>
<td>[0.134; 0.532]</td>
</tr>
<tr>
<td>Indep.</td>
<td>3.474*</td>
<td>7.946***</td>
<td>17.785***</td>
<td>46.437***</td>
<td>50.825***</td>
</tr>
<tr>
<td>F-stat.</td>
<td>13.767</td>
<td>20.234</td>
<td>42.528</td>
<td>23.716</td>
<td>93.358</td>
</tr>
<tr>
<td>Obs.</td>
<td>2,224</td>
<td>2,279</td>
<td>1,873</td>
<td>1,858</td>
<td>8,234</td>
</tr>
</tbody>
</table>

Note. For OLS-ATE and IV-LATE, asterisks indicate statistical significance at the *10%, **5%, and ***1% level. For all estimators, CI are at 5% level. Estimators using an instrumental variable and MTR are run over the same sample; therefore, the number of observations is the same.
the high level of risk aversion, the estimated wage densities cross (Figure 2) and the test of Mourifié and Wan (2016) rejects the null hypothesis most of the time. Therefore, the validity of risk aversion as instrument is falsified in this application.6

An integral part of the analysis is based on the $F$-statistic from the first stage, which is informative concerning the relevance of the instrument. This is reported in Table 2, together with the coefficients attached to the public sector indicator, their confidence intervals (CI) at 95% level, and the related test of uncorrelated errors between the selection mechanism and the wage equation. The $F$-statistic with father’s sector of occupation is always larger than 20 (Stock & Yogo, 2002), with the only exception of 2006. Not shown, the other potential instruments are less relevant to explain the sorting mechanism, as the first step $F$-statistic is small: For example, using the sector of the mother—the second highest $F$-statistic—the highest first-step $F$-statistic is only about 10.

Owing to the critical role played by exclusion restriction, relevance, and monotonicity assumptions, the above analysis fills an important gap in this literature, because it is the first time that the validity of the sector of father as candidate instrument for sorting in the public sector is formally supported on a statistical ground.7

---

6SHIW contains information on other potential instruments for the public sector indicator, namely the parents’ education (see Table 5 in Card, 1999), for a review of studies using family background for the estimation of the return to schooling) and a question about a lottery. The analysis is also conducted using these instruments. They are not discussed within the text to avoid confusion. Indeed, these indicators might not be good instruments satisfying, in particular, the exogeneity assumption, which would be violated if parents’ education affects individual wages (Card, 1999), or if risk aversion is nonconstant-in-wage (Cocco, Gomes, & Maenhout, 2005). Nevertheless, both channels are investigated in a sort of robustness check because, exploiting all available instruments, it is possible to test for overidentifying restrictions ($J$-test; Hansen, 1982) and to possibly eliminate invalid instruments. The $J$-test strongly rejects the null hypothesis that moment conditions are equal to zero, mainly due to the low educational level achieved by the parents and risk aversion, which contribute most to the rejection of the null hypothesis. Based on previous economic concerns, I interpret the result as evidence against these additional instruments rather than as treatment effect heterogeneity.

7Moreover, I cannot reject the null hypothesis that the instrument’s coefficient is zero when using parents’ sector of occupation to explain the mean wage function. On the contrary, for other potential instruments I always reject the hypothesis at standard confidence level: This further reinforces the scepticism about their validity when used as instruments, with these data.
As is usually found in the literature (Bardasi, 1996; Depalo & Giordano, 2011; Dustmann & van Soest, 1998), the estimated gap in favor of the public sector workers using father’s sector of occupation as instrument is of a magnitude that is much larger than under random sampling. Importantly, as this paper clarifies, this point estimate is for compliers’ mean effect: Pooling all the years, it is as large as 33%. Also, the steep increase in the gap estimated between 2010 and 2012—although not statistically significant at standard confidence levels—is in huge contrast to the expectations based on the wage freeze enforced by the Budget Law issued in 2010 (Section 4).

In Section 3 I emphasize that, in general, OLS-ATE and IV-LATE estimate different parameters, the former for the working population and the latter for the subpopulation of compliers. Therefore, in principle OLS-ATE and IV-LATE may be different even if both of them are consistent for their respective parameters. As for IV-LATE, in general, different instruments estimate different marginal returns that may be above or below the OLS-ATE, and each is valid for different subpopulations of compliers. For these reasons, it is important to gain more insight into this return. In the spirit of Card (1999), the IV-LATE parameter can be viewed in terms of a cost–benefit analysis with $\beta = B - C$, where $B$ is individual return due, for example, to motivation, whereas $C$ is the opportunity cost. The couple $\{B, C\}$ defines benefits and costs for each individual, respectively. For simplicity, assume that benefits and costs can take on only two values: high ($H$) or low ($L$). With father’s sector of occupation used as instrument, I identify the pay gap for individuals $\{H, L\}$. Their benefit is high and they are pushed to work in the public sector: Their parents shape their utility and they attach a high value to the public sector. Their cost to access the public sector is low, either because their ability is high (so that their effort to pass the test of admission to the public sector is low) or because they apply for middle/low job positions that are characterized by an easier competition. On the basis of the cost–benefit analysis, the return for these workers should be the highest compared to other possible combinations, where either the benefit is low or the cost is high.

In order to check this interpretation, one would like to individually identify the compliers. This is not possible. However, the distribution of their characteristics can be described (Angrist, 2004; Angrist & Pischke, 2008). Compliers are more likely to be highly educated rather than with secondary education or lower (the likelihoods are 0.666 and 0.938, respectively); white-collar workers rather than blue-collar workers (the likelihood is 0.274), whereas the odds in favor of being a manager are smaller than 1 but remarkably high (0.715); working full time (the likelihood of being part-time is 0.281); marital status seems to play no major role on the odds ratio; by geographical area, there is no clear-cut regularity. Overall, these characteristics are coherent with the economic rationale beyond compliers defined by $\{H, L\}$, possibly thanks to high ability. These individuals are likely to have a large spectrum of opportunities; hence this subpopulation is rather small (as often happens; see Table 4.4.2 in Angrist & Pischke, (2008), and the related discussion): less than 20%. Therefore, unless the circumstances under which LATE and ATE coincide are verified, not specifying which parameter is identified in this application is tantamount to implicitly attributing the parameter that is valid for less than 20% of the
population to the remaining part. To inform on the population average treatment effect, other solutions must be explored. This argument is the strongest motivation for a different approach, which I exploit in the next section.

5.2 Yet another approach: Bounds

To identify the population average treatment effect, I abandon point estimates in favor of bounds (in Table 2 the title of each panel informs us of the technique; see also Figure 3). I begin with bounds under the “monotonic treatment response” (MTR); then I introduce the instrument and estimate bounds under “mean independence of $Z$”, the bounds in Shaikh and Vytlacil (2011), and the narrowest bounds in Bhattacharya et al. (2012). For bounds where I need a finite support $[k_0, k_1]$, I run in-sample statistics and set $k_0$ and $k_1$ equal to the minimum and the maximum observed in the data, respectively. Standard errors are obtained as in Imbens and Manski (2004).

With bounds, the conditioning set is implicitly limited to the time period. However, Bhattacharya et al. (2008, 2012) explicitly mention that if $X$ is contained in $Z$ all of the analysis can be carried out conditional on $X$: I do this in Appendix A3. To the best of my knowledge, this is the standard approach followed in the papers that use bounds (e.g., those mentioned in Section 3) and in particular by Bhattacharya et al. (2008, 2012).

5.2.1 Monotonic treatment response (MTR) bounds

Although the theoretical literature on the public/private wage gap is scant, all existing papers on developed economies predict that the public sector wage cannot be lower than the private sector’s wage (Lausêv, 2014). The reasons are related to the peculiarity of the public employer’s objective function, which includes lobbying (Gunderson, 1978), electoral motives (Fogel & Lewin, 1974), or the exploitation by unions of the relatively inelastic labor demand curve in the public sector (Forni & Giordano, 2003). For this reason, I begin by imposing the MTR such that $Y_1 \geq Y_0, \forall i$. The lower bounds implied by the monotone response assumption are equal to 0 (Manski, 1997). The upper bounds on the population average treatment effect are remarkably high (1.045 in the pooled sample) and therefore the widths of these bounds are very large. Furthermore, as emphasized in Bhattacharya et al. (2012), with MTR assumption one knows a priori that $Y_1 \geq Y_0$ for all individuals; thus the assumption imposes the answer to the question. This is a motivation to adopt a different strategy.

---

8The approach computes confidence intervals that asymptotically cover the true parameter with a fixed probability. Technical details can be found in Imbens and Manski (2004). The underlying idea is that, if the true parameter value is close to the upper bound of the identification region, the asymptotic probability that the estimate for the lower bound exceeds the true value can be ignored when making inference, and the entire probability of making an error $1 - \alpha$ can be allocated to values above the upper-bound point estimate. Since I do not know whether the true parameter is close to the upper or lower bound, one-sided intervals with confidence level $\alpha$ are constructed around both bounds. The CI are equal to $C(\alpha) = [\hat{LB} - C_n \frac{\sigma_{LB}}{\sqrt{n}}, \hat{UB} + C_n \frac{\sigma_{UB}}{\sqrt{n}}]$, where: LB stands for lower bound; UB stands for upper bound; $\sigma_{LB}$ and $\sigma_{UB}$ are bootstrap estimates of the asymptotic standard errors for the estimated lower and upper bounds based on 500 draws; n is the sample size; $C_n$ satisfies $\phi \left[ C_n + \sqrt{n} \frac{(\hat{LB} - \theta_{LB})}{\sigma_{LB}} \right] - \Phi(-C_n) = \alpha$, with $\Phi(\cdot)$ the cumulative distribution function of the standard normal distribution. This part of the program is taken from McCarthy, Millimet, and Roy (2015).
5.2.2 | Mean independence of Z bounds

If one imposes the independence of the father's sector of occupation, the population average treatment effect is consistent with wide bounds, estimating a huge disadvantage and a huge advantage from working in the public sector. With this approach, the width of the bounds amounts to \((k_1 - k_0)[P(D = 0|Z = 1) + P(D = 1|Z = 0)]\). The more relevant the instrument, the smaller the width. In the application, the width is minimized when the instrument is the sector of the father (equal to 4.760 in the pooled sample), as a consequence of \([P(D = 0|Z = 1) + P(D = 1|Z = 0)] = 0.768\), smaller than with any of the other potential instruments. For comparison, using the sector of the mother as instrument (not shown), that is, the second most relevant instrument, the same component is equal to 0.900, and accordingly the width is much larger (5.585).

5.2.3 | Shaikh and Vytalci (2011) bounds

If data were coherent with more structure than that imposed thus far, the bounds could be narrowed. With this aim, I first explore the rank similarity. Chernozhukov and Hansen (2004) offer an interpretation of the property that, ex ante, the (conditional) rank of wages may be considered to be the same across potential treatments, although ex post the realized rank may be different across the two sectors. Chernozhukov and Hansen (2006) in their Section 2.3 complement this explanation with a formal example based on the Roy (1951) model of Section 2. Consider, for example, the ranking function \(f : R \rightarrow [0, 1]\), \(U_d = f(v + \mu_d)\), where \(\mu_d\) is a noisy adjustment of ranking across sectors, relative to observationally equivalent workers. This formulation is very useful because it shows that imposing rank similarity might not be innocuous to the extent that \(\mu_d\) presents some systematic patterns. Instead, if \(\mu_d\) is (conditionally) i.i.d., the optimization problem described in Section 2 is not affected and rank similarity holds. In the latter case, the correlation coefficient of wages in the two sectors conditional on relevant characteristics would be high for all quantiles. Accordingly, I built cells based on combinations of variables belonging to the classical Mincer equation (rather than on the whole set used in the OLS, in order to have a larger number of observations within each cell) and, by decile, the parameter is indeed remarkably high (usually, 0.8 or more).\(^9\) To further support the assumption, following a recent paper by Frandsen and Lefgren (2017), I estimate the ranking function based on potential outcomes \((U_i)\) and run the regression \(U_i = \alpha_0 + \alpha_1D_i + \alpha_2[X_i, Z_i] + \delta D_i[X_i, Z_i] + \epsilon_i\). If rank similarity holds, then \(\delta = 0\): Coherent with the results just described, I could not reject this null hypothesis (even when \(X\) is limited to time dummies only). Based on this evidence, I estimate bounds imposing also rank similarity.

The following bounds are obtained invoking (i) threshold crossing model, (ii) the independence of \(Z\), (iii) the monotonicity (not rejected by the test in Mourifié and Wan (2016), whose results are similar to the testable implications derived by Shaikh and Vytalci (2011) in Remark 2.3; see panel “No control” in Table 3), and (iv) the rank similarity.

The gap estimated using the method in Shaikh and Vytalci (2011) is much more informative than from previous bounds. The average public/private wage gap for the population is positive in all of the years. Since in the analysis \(E[Y|Z = 1] > E[Y|Z = 0]\), it follows that the upper bound with this approach is the same as in the independence bounds, I discuss therefore only the lower bound. Over time, it is 15% or more; when I pool all the waves, the lower bound of the wage gap is 18%. The corresponding confidence intervals are always able to statistically reject a zero population average treatment effect.

To better appreciate the gain with respect to the previous bounds, it is worth looking at how much these bounds narrow and why they do so. Going from bounds under the mean independence assumption to \(SV\), the lower bound for the pooled sample goes from −2.293 to 0.182, a difference equal to 2.475. The reduction is attributable to some conceptually distinct components: One is the “assumption effect” versus the “population effect”; the other is the “never-takers effect” versus the “always-takers effect”:

\[
\Delta_{M-SV}^{LB} = \begin{cases} 
E[Y|D = 0, Z = 1] - k_0 & \text{Never-takers} \\
\{k_1 - E[Y|D = 1, Z = 0]\} & \text{Always-takers} 
\end{cases}
\]

This representation clarifies the contribution of each component. The contribution to \(\Delta_{M-SV}^{LB}\) of each assumption is larger, the larger the population to which it applies; and similarly the contribution of each subpopulation is larger, the larger

\(^9\) Having fewer cells than those in accordance with the whole conditioning set is conservative, in the sense that the richer the set of covariates, the more plausible the rank similarity (Chernozhukov & Hansen, 2006).
the gain from the corresponding assumption.\textsuperscript{10} The economic intuition of the gain can be seen focusing, for example, on never-takers: These workers are not observed in the public sector \((D = 0)\), unlike their fathers \((Z = 1)\), but thanks to the invoked assumptions I know that “imputing” \(k_0\) to their wage in the unobserved status of public sector employment would be “too pessimistic.”

The increase in the lower bound due to never-takers \((1.951, about 80\% of the total gain) is much higher than that due to always-takers (equal to 0.522, the remaining 20\% of the gain), and the increase due to the “assumption effect” is substantial because the assumptions mitigate the heterogeneity in wages, more for never-takers than for always-takers. The population effect further reinforces the gain for the former group.

Equally relevant information regards the width of the bounds. It amounts to

\[
\text{Width}_{SV} = \{k_1 - E[Y|D = 0, Z = 1]\} P(D = 0|Z = 1) + \{E[Y|D = 1, Z = 0] - k_0\} P(D = 1|Z = 0) \]

\text{Never-takers} \quad \text{Population effect}

\text{Assumptions effect} \quad \text{Always-takers.}

Since \(k_1 - E[Y|D = 0, Z = 1] = 2.779\) and \(E[Y|D = 1, Z = 0] - k_0 = 3.549\), the width is 2.283 with a contribution of never-takers \((1.584, or 70\% of the width) as large as twice the contribution of always-takers \((0.699)\).

Even though the rank similarity shrinks the bounds by half with respect to the independence assumption alone, still large heterogeneity persists, overall, for never-takers. At the same time, the fact that a large contribution to \(\Delta_{LB}^{SV}\) and \(\text{Width}_{SV}\) comes from never-takers suggests that much of the heterogeneity in the wage gap comes from that subpopulation.

Given that SV bounds (and BSV bounds presented below) partially identify the \textit{population} average treatment effect, while IV-LATE (in general) point identifies the effect for \textit{compliers} only, to conclude the inspection of SV bounds I compare the two methods. The lower bounds for the population average treatment effect are 1.5–2 times smaller than the corresponding IV-LATE estimates for the compliers effect. I interpret this as evidence that a single number summarizing everything is inappropriate to shed light on the mean generating process of the public/private wage gap, because of the heterogeneity of responses to treatment and because \(Y_1 (Y_0)\) is never observed for never-takers (always-takers). Even though IV-LATE for compliers is still within the SV bounds on the population average treatment effect, the data are compatible also with a much smaller average gap in favor of the public sector workers. At this point, one may be tempted to jump to the conclusion that the random allocation (OLS-ATE) is perfectly consistent with these new results and thus there is no gain from using more sophisticated approaches. This claim would be simplistic and wrong. Indeed, the OLS-ATE estimates are always smaller than the lower bound of the SV method, thus pointing toward a nonignorable sector sorting.

\subsection*{5.2.4 Bhattacharya et al. (2012) bounds}

In an attempt to further shrink the bounds, I explore whether the data are coherent with even more structure added, namely the PQD hypothesis (or first stochastic dominance). To this aim, I checked the cumulative distribution function of wages \((F_D(Y))\) by sector of employment and instrument and verify that \(F_1(Y) \leq F_0(Y)\) for all wage levels, with strict inequality for some levels. More formally, the test statistics of the null hypothesis proposed by Barrett and Donald (2003) of first-order stochastic dominance of the public versus private wages for the pooled sample are always smaller than 0.5, whereas they are larger than 2 for the reverse null hypothesis (i.e., first-order stochastic dominance of the private vs. public wages); these numbers should be compared to the 5\% critical values equal to 1.224 (Barrett and Donald, 2003, p. 78; 1.073 at 10\% and 1.517 at 1\%). By year, results are similar. Therefore, the public sector wage distribution first-order stochastic dominates the private sector’s distribution, but not vice versa. This is indirect empirical evidence in favor of the PQD hypothesis.

Based on this analysis, I impose the following assumptions: (i) the threshold-crossing model; (ii) independence of \(Z\); (iii) monotonicity; (iv) rank similarity; and (v) PQD; and estimate the bounds on the population average treatment effect in Bhattacharya et al. (2012).

Because the lower bound is the same as in SV bounds, I now consider only the upper bound. The upper bound of the gap is in the range 30–45\%, by year; it is about 35\% for the pooled sample. The difference in the upper bound between BSV and SV bounds amounts to

\textsuperscript{10}In Table 2 the relevant quantities are: \(E[Y|D = 0, Z = 1] - k_0 = 3.422, P(D = 0|Z = 1) = 0.570, k_1 - E[Y|D = 1, Z = 0] = 2.652, \) and \(P(D = 1|Z = 0) = 0.197, so that (apart from rounding) 3.422 \times 0.570 + 2.652 \times 0.197 = 2.475.\)
To have an intuition related to this difference: consider again never-takers. Thanks to PQD, I know that the wage of workers not observed in the public sector—unlike their fathers—cannot be higher than the wage of workers observed in the public sector—as with their fathers (i.e. $E[Y|D=1, Z=1]$): imputing $k_1$ to their unobserved status of public sector employment would be “too optimistic.”

The gain from stochastic dominance is substantial as the upper bound from BSV is much smaller than from SV (the difference is 2.108). The never-takers effect is 70% (1.460) of the total $\Delta_{\text{SV-BSV}}$, because the smaller assumption effect for these workers than for always-takers is more than balanced by the population effect.

The width of the BSV bounds is equal to

$$\text{Width}_{\text{BSV}} = \{E[Y|D=1, Z=1] - E[Y|D=0, Z=1]\} P(D=0|Z=1) + \{E[Y|D=1, Z=0] - E[Y|D=0, Z=0]\} P(D=1|Z=0)$$

The contribution to the total width (0.177) of “never-takers” (0.125) is larger than that of “always-takers” (0.053). Although BSV bounds are much narrower than other bounds, they are still wide; thus the analysis confirms that the working population is highly heterogeneous. Also, the larger contribution of “never-takers” confirms and complements the results from SV bounds that the heterogeneity due to this subpopulation is substantial.

At this point, it may be useful to discuss the results from the narrowest bounds on the population average treatment effect. The bounds on the population average treatment effect in favor of men working in the public sector are wide, in the range 18–36% when I pool all the waves. Although a wide range may be unpleasant for the policymaker, it is of great interest to better understand the wage-generating process in the two sectors and as a check for IV-LATE for compliers (Nicoletti, 2010). Some remarks are in order. First, that width remains wide even imposing more structural assumptions confirms that one should be very careful when imposing them and always support them even in a partial identification framework (as Ho and Rosen, (2015), put it: partial identification is not a panacea for using assumptions). Second, if one is willing to impose further assumptions to narrow the bounds, looking at the population of never-takers, where heterogeneity plays a major role, might be a promising direction: of course, whether such restrictions are credible or not should be evaluated case by case, preserving the credibility of the inference (according to the “law of decreasing credibility” in Manski, (2011)). Third, that IV-LATE (for compliers) is usually economically indistinguishable from or above the upper bound (for the population average treatment effect) confirms the need to be very precise about what parameters are actually identified. Since the estimated IV-LATE is the treatment effect for compliers and not for the entire population—except for the special cases listed in Section 3—it could be the case that it is consistent, but outside the bounds on the population average treatment effect. Fourth, at the same time, that OLS-ATE is below the lower bound rules out the possibility of random sector sorting.

Between 2010 and 2012, after the Budgetary Law in 2010 introduced a wage freeze in the public sector, the lower bound remained almost constant (from 18% to 21%), whereas the upper bound greatly increased from 34% to about 43%. The lower bound, which is broadly consistent with expectations, was more affected than the upper bound. Finally, it is noteworthy that the width of bounds increased between the last two waves, thus suggesting a larger heterogeneity in the wage differential between private and public sector workers.

6 | CONCLUSIONS

A gap in favor of the public sector workers is estimated in (almost all) developed countries. Existing studies for Italy are in line with this finding. This paper is the first attempt to shed light on what quantities of the public/private wage gap are identified and under what conditions. Italy is an interesting example to analyze in this context because the public sector

---

11The relevant components are as follows: $k_1 - E[Y|D=1, Z=1] = 2.560$ and $E[Y|D=0, Z=0] - k_0 = 3.282$.

12The numbers follow from $E[Y|D=1, Z=1] = E[Y|D=0, Z=1] = 0.219$ and $E[Y|D=1, Z=0] = E[Y|D=0, Z=0] = 0.267$.

13For example, Chen et al. (2017) consider mean dominance assumptions on the average potential outcomes of never-takers, always-takers, and compliers, some of which result in narrower bounds than those in Bhattacharya et al. (2012) in their application.
employment is sizable, the wage bill is high, the country underwent a serious public finance difficulty during the latest economic crisis that led to a wage freeze during the period 2011–2016, and the dozens of existing papers on this issue estimate a large wage gap in favor of the public sector.

Results from standard estimators are similar to those obtained in the existing literature (see Depalo & Giordano, 2011), for a summary of the existing literature the country). The classical OLS-ATE approach would be consistent with a wage gap in favor of public sector workers equal to 5% with respect to their private sector counterparts; as soon as I allow for the possible sorting using the father's sector of occupation as instrument, the gap for compliers increases to above 30%. A first contribution of the paper is the clarification that this high differential is a local gap for compliers, that is, individuals who are induced to work in the public sector because their fathers worked in the public sector (Imbens & Angrist, 1994). The distribution of their characteristics (Angrist, 2004) is coherent with the economic rationale of this population made of individuals who enjoy the highest return from working in the public sector. Since IV-LATE relies on a suitable instrument (Angrist et al., 1996), a second contribution of the paper is the support for the instrument used in the analysis, on the basis of a formal statistical ground (Mourifié & Wan, 2016) rather than only on economic arguments, as always done so far.

To identify the population average treatment effect with an IV, instead of a local gap for compliers only, other solutions must be investigated. The novelty of this paper for the public/private wage gap literature is to learn about the population average treatment effect with an IV, under various relatively mild assumptions, using partial identification. Using the narrowest bounds on the population average treatment effect of Bhattacharya et al. (2012), I estimate a lower bound that is always higher than the gap that is estimated when imposing random sector sorting. The IV-LATE (for compliers) is economically indistinguishable from or above the upper bound (for the population average treatment effect). Therefore, understanding what parameter is actually identified is key to correctly interpreting the results. The bounds on the population average treatment effect are also consistent with a much smaller gap than that estimated by IV-LATE for compliers. The conclusions that may be drawn from these results are still relevant for policymakers and researchers, as the admissible range of returns always lies on the positive side, between 18% and 36% in the pooled sample. This wide range should not be seen as unpleasant, but simply as the measure of heterogeneity in the workforce (Horowitz & Manski, 2000).

In Italy, of particular interest is the period after 2010, when a public sector wage freeze was introduced. This paper is not an evaluation of that policy, because in only one year little effect was expected. However, between 2010 and 2012 the lower bound of the gap remained relatively constant, whereas the upper bound increased greatly.

ACKNOWLEDGEMENTS
I am sincerely grateful to two anonymous referees, whose comments were very helpful in improving the paper, and to Edward Vytlacil. I am indebted to Monica Andini. I also thank Erich Battistin, Aureo de Paula, Cristina Gualdani, Toru Kitagawa, Ismael Mourifié, Santiago Pereda, Enrico Rettore, Marco Savegnago, and Jeffrey Wooldridge. I benefited from comments received at the Third Annual Conference of the International Association for Applied Econometrics (IAAE) in Milan. Replication files and additional results will be available on the web page, at: http://sites.google.com/site/domdepalo/. The views expressed in this paper are those of the author and do not responsibility of the Bank of Italy.

REFERENCES


How to cite this article: Depalo D. Identification issues in the Public/Private wage gap with an application to Italy. J Appl Econ. 2018;33:435–456. https://doi.org/10.1002/jae.2608
APPENDIX A: THE PUBLIC SECTOR IN ITALY AND RELATED LITERATURE

In this appendix I describe the public sector in Italy and the related literature on the public/private wage gap. A fairly detailed description of the Italian public sector, along with a comparison with France and the UK, is given in Lucifora and Meurs (2006); a comparison with Germany is given in Brunello and Dustmann (1997); a comparison with other EU countries on a legal basis is given in ILO Staff (2015). Here, I outline the main characteristics of the public sector in Italy and refer the reader to the original papers for further details.

As in many European countries, in Italy the employment and pay conditions between the public and private sectors are quite different, especially with respect to (1) the criteria adopted to select, recruit, and promote workers and (2) the role played by collective bargaining and trade unions during the wage setting. The recruitment in the public sector is through an open, competitive, anonymous examination for which requisites and rules are published in the Italian Official Journal (Gazzetta Ufficiale), and, once hired, workers enjoy lifetime contracts in which seniority plays a major role in terms of wage and career (Lucifora & Meurs, 2006). As a general rule, public servants cannot be fired, except for misconduct, and the statutory terms apply regardless of whether the individual is employed at the national, regional, or local authority level.14

Public sector wages (levels and growth) have been largely decided at the central level since 1993, when a reform assigned a larger role to collective bargaining and created an independent agency (Agenzia per la Rappresentanza Negoziale nella Pubblica Amministrazione—ARAN) that is responsible for negotiating pay levels and working conditions for most public sector employees. The goal of the reform was to make the determinants of payment and employment conditions in the public sector closer to those in the private sector, through a greater role for negotiation, tighter constraints on wage growth, and the replacement of automatic bonuses with schemes based on merit. It is a shared opinion that the reform failed to achieve its main targets (e.g., Lucifora, 1999), so that the public sector is currently undergoing a new reform (Enabling Law 124/2015, Decision 251/2016 of the Constitutional Court and the recent Decree Law (http://www.funzionepubblica.gov.it/i-decreti-attuativi). If the new measures are structural and well designed, they might improve the efficiency of the public sector and support the public wage moderation through several channels, such as competitiveness and complementarity with the private sector (Pèrez et al. 2016).

A.1 Related literature

An updated review of the literature on the public/private wage gap is presented in Giordano et al. (2014) and Lausèv (2014). A focus on Italy can be found in Depalo and Giordano (2011).15 Most of the early literature in this field is based on US data and imposes ignorable selection such that a consistent estimate of the average gap is obtained using an OLS on a Mincerian equation (Smith, 1976); during the last 10–15 years, increasing interest has been devoted to European countries. Evidence for Italy points towards a pay gap in favor of public sector male workers of about 5–10% (e.g., Brunello & Dustmann, 1997). Considering the sorting mechanism is standard, nowadays. A complete treatment of selection mechanism can be found in Dustmann and van Soest (1998), where the sector decision is modeled for men in Germany, together with schooling decision. After the correction is made, a significantly higher pay gap than under ignorable selection is estimated. For Italy, Bardasi (1996) estimates a wage advantage in the public sector of up to 35% after correction; Depalo and Giordano (2011) estimate a pay gap higher than 30%.

APPENDIX B: THE DATA

The data used in the analysis are from the Survey on Household Income and 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 (age, gender, marital status, educational level, region of residency) and occupational (sector of

---

14 Rules have changed for the private sector in 2014 with the approval of the so-called Jobs Act, which introduced new regulations lowering firing costs and making them less uncertain (Sestito & Viviano, 2016). The original idea of (at least part of) the Government was to extend the Jobs Act to the public sector, but a recent decision of a court (Corte di Cassazione, Decision 11868/2016) clarified that the new private sector’s rules as such do not apply directly to the public sector (see below).

15 In this review, I consider only the average wage gap. An interesting departure (not considered here) is related to quantiles: Examples with a focus on Italy include Lucifora and Meurs (2006) and Depalo and Giordano (2011).
economic activity, occupational level, firm size, part-time status, number of months worked in the year, average number of hours worked in a week) characteristics, wages (net of income and payroll taxes) and type of activity.

Three aspects that deserve attention for the analysis presented in this paper are: the definition of the public sector; the definition of wage; and the existence of variables that may be used as appropriate instruments.

Following Lucifora and Meurs (2006), I define someone as a public sector employee if his sector of activity is “public administration, defence, education, health and other public services.” This definition is consistent with the figures from Italian National Accounts and allows a time series dimension. Nevertheless, it may induce some measurement errors, for example, because workers of the health sector would be defined as public sector employees even though they work for private providers.16 Public sector employment in Italy is sizable, with about 3.3 million workers in 2012, which is a reduction from 3.6 million workers in 2006. From National Accounts, the largest drop in the units of labour was recorded in 2010 and 2012, a trend that is found also in SHIW. According to this trend, the share of public sector male workers in total employment decreased from about 15% in 2006 to 11% in 2012.

Having defined the public sector, a key aspect is related to the choice of the appropriate variable for wage comparison. Indeed, as the average number of hours worked in a week in the public sector (36) is lower than that reported by workers in the private sector (40), using the monthly wage could underestimate the wage differential across sectors. Thus, in the benchmark specification, I approximate the hourly wage as \( \frac{Y W}{M} \div (4 \times \bar{H}) \), where \( Y W \) is yearly wage, \( M \) is number of months worked in the year, and \( \bar{H} \) is average number of hours worked in a week. In Appendix A3, all of the results are checked by also using the monthly definition.

An advantage of SHIW over other datasets is the richness of variables that may be used as instruments. Appropriate instruments influence the decision to work in the public sector, but not wage. The existing literature emphasizes the importance of motivation, ability, and risk aversion. With this aim, I investigate various possibilities, namely the sector of parents, the education of parents, and a specific question about risk aversion.17 When using family background indicators, the definitions of the variables for individuals and their parents is identical.

### B.1 | Descriptive statistics

Descriptive statistics refer to employees aged 20–65 years, which I conventionally define as the working age population.

In Table B1, I report some relevant descriptive statistics for individual characteristics of the sample considered in this paper, pooling all the waves of the sample. Unconditionally, there exists a wage gap in favor of the public sector. When the focus is on hourly wage the gap is 25%, but when it is on monthly wage the gap is scaled down by 10 points, reflecting that public sector workers on average work almost 5 hours less than private sector workers. At the same time, the variance of the wage distribution between the two sectors is roughly the same. I also investigated the minimum and the maximum wage by sector, because the bounds are usually larger when \([k_0, k_1]\) are further away: This measure of heterogeneity is larger with hourly wages than with monthly wages.

Public sector workers are older than private sector workers and they are better educated (lower education is achieved in more than 50% of cases in the private sector and slightly less than 30% in the public sector). Notable differences are recorded also by job positions: In the private sector, the percentage of blue-collar workers is as large as twice that of the public sector; a huge difference is observed also for the managerial position, whose incidence is larger by about 15 percentage points in the public sector than in the private sector; however, the greatest difference across the two sectors is in the percentage of white-collar workers, where the public/private proportion is 5 to 1.

As for variables that could potentially be used as instruments, one out of four of the public sector workers has a father who worked in the public sector at the current age of the interviewed survey participants, as opposed to those working in

---

16 For the empirical analysis presented in Section 5, this misreporting would lead to a conservative estimation of the differential. Besides the classical measurement error analytics (i.e., an attenuation bias), if public sector workers enjoy a wage advantage—as expected, based on the National Accounts data—the classification of some private sector workers as public would depress the average wage of the latter group, pushing the wage gap towards zero.

17 One question in the survey is “In managing your financial investments, would you say you have a preference for investments that offer: (a) very high returns, but with a high risk of losing part of the capital; (b) a good return, but also a fair degree of protection for the invested capital; (c) a fair return, with a good degree of protection for the invested capital; (d) low returns, with no risk of losing the invested capital?”
TABLE B1 Summary statistics.

<table>
<thead>
<tr>
<th>Var.</th>
<th>Private Mean</th>
<th>10th</th>
<th>50th</th>
<th>90</th>
<th>Public Mean</th>
<th>10th</th>
<th>50th</th>
<th>90</th>
<th>Test (P-value)</th>
<th>Diff. in means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Hourly wage)</td>
<td>2.18</td>
<td>0.41</td>
<td>1.75</td>
<td>2.14</td>
<td>2.63</td>
<td>2.45</td>
<td>0.42</td>
<td>2.02</td>
<td>2.39</td>
<td>3.00</td>
</tr>
<tr>
<td>ln(Monthly wage)</td>
<td>7.24</td>
<td>0.40</td>
<td>6.82</td>
<td>7.22</td>
<td>7.70</td>
<td>7.41</td>
<td>0.41</td>
<td>6.99</td>
<td>7.39</td>
<td>7.82</td>
</tr>
<tr>
<td>Adjusting variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>42.33</td>
<td>12.73</td>
<td>25.00</td>
<td>42.00</td>
<td>60.00</td>
<td>46.19</td>
<td>9.79</td>
<td>33.00</td>
<td>47.00</td>
<td>58.00</td>
</tr>
<tr>
<td>Low ed.</td>
<td>0.54</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.27</td>
<td>0.44</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Middle ed.</td>
<td>0.36</td>
<td>0.48</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.43</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>High ed.</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.30</td>
<td>0.46</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Blue collar</td>
<td>0.35</td>
<td>0.48</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.14</td>
<td>0.35</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>White collar</td>
<td>0.13</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.62</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Manager</td>
<td>0.04</td>
<td>0.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.18</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Married</td>
<td>0.61</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.75</td>
<td>0.43</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Candidate Instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother pub.</td>
<td>0.08</td>
<td>0.27</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Father pub.</td>
<td>0.12</td>
<td>0.33</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.27</td>
<td>0.44</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Mother high ed.</td>
<td>0.16</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.17</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Father high ed.</td>
<td>0.17</td>
<td>0.38</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.23</td>
<td>0.42</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Mother low ed.</td>
<td>0.61</td>
<td>0.49</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.64</td>
<td>0.48</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Father low ed.</td>
<td>0.57</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.53</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Low risk</td>
<td>0.51</td>
<td>0.50</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.43</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>High risk</td>
<td>0.16</td>
<td>0.36</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.18</td>
<td>0.39</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

the private sector, for whom the share is one in 10. Similarly, for public sector workers the share of highly educated parents, namely the father, is higher than in the private sector, although by a small margin. Finally, unconditional statistics show that individuals accepting only low risk are more in the private sector than in the public sector.

APPENDIX C: ROBUSTNESS CHECKS

In this appendix, I run robustness checks to test whether results are driven by low wage levels or by the focus on hourly rather than on monthly wage definition (Table C1). I also condition on age, a key covariate for the public/private wage gap in Italy (Appendix A1 and Lucifora and Meurs, 2006), following the approach advocated in Bhattacharya et al. (2008, 2012).

In the spirit of a robustness check, I report only bounds from Bhattacharya et al. (2012). Other bounds are estimated but not shown; they are available on my web site.

C.1 Minimum wage

In Italy, there is no legal obligation for imposing a minimum wage; therefore, the benchmark analysis is carried over the entire sample. To the best of my knowledge, in the literature there is no exception to this practice. Nonetheless, very low hourly wages do not seem entirely credible: It is important to check whether they drive the results. For this reason, as a robustness check I drop observations below the threshold of 5 euros per hour (about 1.5% of the sample in the public sector and about 2.5% in the private sector). This amount is comparable to those set by law in other European countries that underwent a serious public finance tension during the last economic downturn—for example, Greece, Portugal, or Spain—but slightly lower than in Belgium, France or Germany. I view this threshold as a fair compromise between the lack of a binding legislation on minimum wage in Italy and the need for dropping very low hourly wages.

Since wages below the threshold are encountered more in the private than in the public sector, the wage gap decreases, but only marginally (when significant). This justifies the standard approach of the literature. Also, all previous findings regarding the OLS-ATE for the population below the lower BSV bound on the population average treatment effect, and the IV-LATE for compliers that is indistinguishable from or slightly above the upper bound, are confirmed.
### TABLE C1  Robustness check

<table>
<thead>
<tr>
<th>Gap</th>
<th>2006</th>
<th>2008</th>
<th>2010</th>
<th>2012</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Minimum wage</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OLS-ATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.034**</td>
<td>0.057***</td>
<td>0.032**</td>
<td>0.066***</td>
<td>0.046***</td>
</tr>
<tr>
<td>CI</td>
<td>[0.004; 0.064]</td>
<td>[0.026; 0.087]</td>
<td>[0.000; 0.064]</td>
<td>[0.033; 0.098]</td>
<td>[0.031; 0.062]</td>
</tr>
<tr>
<td><em>IV-LATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.254</td>
<td>0.462**</td>
<td>0.339**</td>
<td>0.395**</td>
<td>0.369***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.201; 0.708]</td>
<td>[0.073; 0.851]</td>
<td>[0.064; 0.615]</td>
<td>[0.019; 0.771]</td>
<td>[0.187; 0.552]</td>
</tr>
<tr>
<td>Indep.</td>
<td>1.050</td>
<td>6.621***</td>
<td>65.805***</td>
<td>51.417***</td>
<td>53.732***</td>
</tr>
<tr>
<td>F-stat.</td>
<td>13.416</td>
<td>22.882</td>
<td>41.772</td>
<td>23.706</td>
<td>95.406</td>
</tr>
<tr>
<td><em>Bhattacharya et al. (2012)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.146</td>
<td>0.183</td>
<td>0.186</td>
<td>0.196</td>
<td>0.178</td>
</tr>
<tr>
<td>Upper</td>
<td>0.301</td>
<td>0.346</td>
<td>0.305</td>
<td>0.390</td>
<td>0.337</td>
</tr>
<tr>
<td>CI</td>
<td>[0.104; 0.361]</td>
<td>[0.134; 0.412]</td>
<td>[0.141; 0.361]</td>
<td>[0.145; 0.457]</td>
<td>[0.153; 0.370]</td>
</tr>
<tr>
<td><strong>Monthly wage definition</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OLS-ATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.034**</td>
<td>0.057***</td>
<td>0.032**</td>
<td>0.066***</td>
<td>0.046***</td>
</tr>
<tr>
<td>CI</td>
<td>[0.004; 0.064]</td>
<td>[0.026; 0.087]</td>
<td>[0.000; 0.064]</td>
<td>[0.033; 0.098]</td>
<td>[0.031; 0.062]</td>
</tr>
<tr>
<td><em>IV-LATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.254</td>
<td>0.462**</td>
<td>0.339**</td>
<td>0.395**</td>
<td>0.369***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.201; 0.708]</td>
<td>[0.073; 0.851]</td>
<td>[0.064; 0.615]</td>
<td>[0.019; 0.771]</td>
<td>[0.187; 0.552]</td>
</tr>
<tr>
<td>Indep.</td>
<td>3.250*</td>
<td>3.954**</td>
<td>5.634**</td>
<td>5.485**</td>
<td>17.241***</td>
</tr>
<tr>
<td>F-stat.</td>
<td>11.955</td>
<td>19.881</td>
<td>39.354</td>
<td>20.417</td>
<td>84.852</td>
</tr>
<tr>
<td><em>Bhattacharya et al. (2012)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.122</td>
<td>0.168</td>
<td>0.142</td>
<td>0.189</td>
<td>0.155</td>
</tr>
<tr>
<td>Upper</td>
<td>0.178</td>
<td>0.288</td>
<td>0.261</td>
<td>0.324</td>
<td>0.261</td>
</tr>
<tr>
<td>CI</td>
<td>[0.072; 0.258]</td>
<td>[0.124; 0.347]</td>
<td>[0.092; 0.321]</td>
<td>[0.136; 0.389]</td>
<td>[0.132; 0.294]</td>
</tr>
<tr>
<td><strong>Workers older than 40 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OLS-ATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>−0.007</td>
<td>0.023</td>
<td>0.037*</td>
<td>0.067***</td>
<td>0.029***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.049; 0.036]</td>
<td>[−0.017; 0.063]</td>
<td>[−0.004; 0.077]</td>
<td>[0.024; 0.109]</td>
<td>[0.009; 0.050]</td>
</tr>
<tr>
<td><em>IV-LATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-stat.</td>
<td>3.327</td>
<td>2.858</td>
<td>23.656</td>
<td>11.571</td>
<td>34.262</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.659; 1.370]</td>
<td>[−0.519; 2.175]</td>
<td>[−0.398; 0.266]</td>
<td>[−0.186; 0.866]</td>
<td>[−0.025; 0.570]</td>
</tr>
<tr>
<td>Indep.</td>
<td>7.987***</td>
<td>5.079**</td>
<td>0.053</td>
<td>12.338***</td>
<td>12.982***</td>
</tr>
<tr>
<td>F-stat.</td>
<td>3.327</td>
<td>2.858</td>
<td>23.656</td>
<td>11.571</td>
<td>34.262</td>
</tr>
<tr>
<td><em>Bhattacharya et al. (2012)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.161</td>
<td>0.229</td>
<td>0.128</td>
<td>0.198</td>
<td>0.182</td>
</tr>
<tr>
<td>Upper</td>
<td>0.251</td>
<td>0.361</td>
<td>0.259</td>
<td>0.374</td>
<td>0.314</td>
</tr>
<tr>
<td>CI</td>
<td>[0.093; 0.356]</td>
<td>[0.164; 0.447]</td>
<td>[0.068; 0.330]</td>
<td>[0.139; 0.453]</td>
<td>[0.152; 0.356]</td>
</tr>
<tr>
<td><strong>Workers younger than 40 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>OLS-ATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>0.055*</td>
<td>0.121***</td>
<td>0.035</td>
<td>0.077**</td>
<td>0.072***</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.002; 0.112]</td>
<td>[0.065; 0.178]</td>
<td>[−0.031; 0.101]</td>
<td>[0.004; 0.149]</td>
<td>[0.041; 0.103]</td>
</tr>
<tr>
<td><em>IV-LATE</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wage gap</td>
<td>−0.015</td>
<td>0.095</td>
<td>0.762**</td>
<td>0.405</td>
<td>0.272*</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.654; 0.625]</td>
<td>[−0.318; 0.508]</td>
<td>[0.083; 1.440]</td>
<td>[−0.386; 1.197]</td>
<td>[−0.017; 0.562]</td>
</tr>
<tr>
<td>Indep.</td>
<td>0.091</td>
<td>0.024</td>
<td>21.407***</td>
<td>17.435***</td>
<td>5.013**</td>
</tr>
<tr>
<td><em>Bhattacharya et al. (2012)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower</td>
<td>0.086</td>
<td>0.099</td>
<td>0.184</td>
<td>0.088</td>
<td>0.109</td>
</tr>
<tr>
<td>Upper</td>
<td>0.213</td>
<td>0.245</td>
<td>0.340</td>
<td>0.283</td>
<td>0.258</td>
</tr>
<tr>
<td>CI</td>
<td>[−0.054; 0.278]</td>
<td>[0.036; 0.318]</td>
<td>[0.074; 0.391]</td>
<td>[−0.162; 0.358]</td>
<td>[0.068; 0.312]</td>
</tr>
</tbody>
</table>

Note. For OLS-ATE and IV-LATE, asterisks indicate statistical significance at the *10%, **5%, and ***1% level. For all estimators, CI are at 5% level.
The empirical content of this robustness check is greatly enriched if I consider bounds alone. Pooling all the waves, the independence bounds with minimum wage (not shown) are \([-0.769; 1.888]\), which compare to those from the benchmark analysis that are equal to \([-2.293; 2.467]\). The gain in terms of smaller width is substantial (by half or so) because now \(k_0\) is higher. Also, the width of BSV bounds decreases from the benchmark analysis to the minimum wage (on average by 2 percentage points), as a consequence of (1) the difference in the lower bounds between the two cases being economically irrelevant, while (2) the difference for the upper bounds is larger. This behavior suggests that the upper bound of the gap may be due to low wage level. Therefore, while confirming previous results, this robustness check brings more information into the analysis.

C.2  |  Monthly wage

If working hours is not a variable of choice for the workers, monthly earning better represents the key remuneration for employees. As for monthly wage definition, under random sampling the gap is smaller by 5–8 percentage points than with hourly definition: Public sector workers are no longer at an earning advantage (in 2006 there was a statistically significant disadvantage of 4%). These point estimates are downward biased because they do not consider the possible selection of the sample: Correcting for the selection mechanism, in pooling all the waves the gap for compliers becomes 25%—about 8.5 percentage points less than the corresponding estimate with hourly wage definition. As in the benchmark case, the IV-LATE for compliers—when significant—is economically indistinguishable from, or larger than, the upper bound on the population average treatment effect.

From such bounds as in Bhattacharya et al. (2012), several aspects are worth emphasizing. Firstly, on average workers still enjoy a higher wage in the public sector than in the private sector as lower bounds are always positive at standard confidence levels. The width of the bounds is much smaller using monthly rather than hourly definition (almost 10 percentage points). Also, the difference between the two wage definitions is much larger at the upper bound than at the lower bound (usually by 7–8 percentage points). An insight into the larger correction at the upper bound is obtained by looking at the distribution of hours worked. The difference in hours worked in the public and private sectors at low quantiles of the hours distribution is larger than the difference at high quantiles. Since, at low quantiles of the hour distribution, the wage gap is higher than at high quantiles (i.e., it is the upper bound of the gap), considering monthly wage annihilates the wage difference and the larger correction at the upper bound follows as a consequence. 18

By and large, this robustness check confirms all of the refinements over the existing literature that were found in Sections 5.1–5.2.

C.3  |  Age

It has been argued that the Italian labour market is segmented; therefore, workers of younger ages are at a wage disadvantage with respect to workers of older ages. By conditioning on this characteristic (less/more than 40 years of age) with the strategy suggested by Bhattacharya et al. (2008, 2012), it is possible to check a possible form of dualism that may hurt the younger workers. Based on the OLS-ATE, in the pooled sample the hourly wage gap is higher for the younger than for the older generations by about 4.5 percentage points; this difference is almost symmetric about the benchmark (2 more points for the younger population; 2.5 less for the older population). When I allow for possible sorting in the public sector, the wage gap in favor of compliers (workers who are induced to work in the public sector because their fathers worked in the public sector) is about 27% for both younger and older generations, smaller than in the benchmark specification by 6 percentage points.

Moving to bounds, in general the OLS-ATE for the population is below the lower bound of BSV on the population average treatment effect, whereas IV-LATE for compliers—when significant—is higher than or close to the upper bounds. This robustness check confirms the benchmark analysis.

Using bounds, both the lower and the upper bounds on the average wage gap are higher for the older workers than for the younger workers. The difference between the two groups is always similar between the lower and the upper bounds, and larger than 5 percentage points in the pooled sample.

18 An example may help clarify this point. Let the monthly wages in the public and private sectors be the same \(w = w_G = w_p\), whereas the hours are \(H_G = H\) in the public sector and \(H_p = H + \epsilon\) in the private sector at low quantiles; \(H_G = H + \epsilon\) and \(H_p = H + 2\epsilon\) at high quantiles. It follows that the bounds are \([w_G; w_p] = [w_H + \epsilon; w_H + 2\epsilon]\) with hourly definition, because \(\frac{w_H + \epsilon}{w_H + 2\epsilon} > \frac{w_H}{w_H + \epsilon}\); whereas \([0; 0]\) with monthly definition, by construction. As a consequence, the correction is larger for upper bound than for lower bound.