WORKING PAPER NO. 389

Estimating Labor Demand Function in the Presence of Undeclared Labour: A Look Behind the Curtain

Edoardo Di Porto and Leandro Elia

February 2015

University of Naples Federico II　　University of Salerno　　Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS – UNIVERSITY OF NAPLES
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unisa.it

WORKING PAPER NO. 389

Estimating Labor Demand Function in the Presence of Undeclared Labour: A Look Behind the Curtain

Edoardo Di Porto[*] and Leandro Elia[**]

Abstract
This paper presents estimates of the own-wage elasticity for undeclared labour demand and calculates the e↲ects of undeclared work on declared wages of various skill levels. To identify the parameters of interest, we exploit a quasiexperimental setting created by three tax amnesty laws brought brought into force in 2002 in Italy. Our main results indicate that an upward shift in undeclared work decreases undeclared wages, increases declared wages, and reduces wage inequality in the declared sector. We find q-complementarity between undeclared workers and low to medium-skilled workers.

Acknowledgements: A previous version of this paper is circulated with the title "Undeclared Work and Wage Inequality". We are grateful to James Alm, Nicola Persico, Henry Ohlsson, Michele Bernasconi, Kim Byung-Yeon, Aloys Pynz, David L. Roth, Frédéric Jouneau-Sion, Nicolas Sahuguet, Helen Garrett, Luisa Araujo as well as the participants at the Shadow conference in Müenster and the seminar participants at the EQUIPPE USTL/Lille for useful comments and suggestions. All errors are our own.

[*] University of Naples Federico II, CSEF and UCFS, Uppsala University

[**] Centre for Research on Impact Evaluation (CRIE), Econometrics and Applied Statistics Unit, European Commission - DG Joint Research Centre. Corresponding author: Centre for Research on Impact Evaluation (CRIE), Econometrics and Applied Statistics Unit, European Commission - DG Joint Research Centre, Via E. Fermi 2749, TP 361, 21027 ISPRA (VA), ITALY, tel.: +39 0332 783077, email: leandro.elia@jrc.ec.europa.eu. The views expressed are purely those of the writer and may not in any circumstances be regarded as stating an official position of the European Commission.

Table of contents

1. Introduction

This paper presents estimates of the own-wage elasticity for undeclared labour demand and calculates the effects of undeclared work on declared wages of various skill levels. Undeclared work is defined as any paid lawful activities hidden to public authorities in order to avoid paying payroll taxes.[1] Understanding the interplay between formal and informal sector is extremely important since in many economies, there is substantial economic activity in the informal sector. This is true for both developing and developed economies.[2] According to Jütting & Laiglesia (2009) out of a global working population of 3 billion workers, nearly two-thirds (1.8 billion) are undeclared or informal workers. Maloney (2004) have found that 30 to 70 percent of urban workers in Latin America are employed in the informal sector. Schneider & Enste (2000) have estimated that in Europe the size of the unofficial economy accounts for 10 to 30 percent of GDP, with Southern European countries and Central European transition economies exhibiting the highest figures.[3]

The relative shortage of convincing studies on this topic is mainly due to the difficulty in collecting data on workers who illegally evade labour taxes. In this paper, we are able to spot those workers thanks to information on evasion of social security contributions included in the Bank of Italy's Survey on Household Income and Wealth.[4] In particular, individuals are asked to report whether employer have paid social security contributions on their behalf. Combining this information with reported earnings we are subsequently able to build an individual indicator of informality.

The paucity of empirical investigations is also the result of the complexity of the phenomenon: the dynamics of informal labour are driven by factors related to both supply and demand. Participation in the informal sector is motivated by a number of reasons including changes in tastes, inherited culture and a lack of access to formal jobs. But workers also participate in the informal sector because there a demand for their labor services exists, which are generally associated with lower labour costs and more flexibility in job termination.

To explore how undeclared work affects undeclared and declared wages, we require an exogenous variation in the supply of undeclared work. In this paper, to identify labour demand elasticities for undeclared and declared work, we have exploited a quasi-experimental setting, created by three tax amnesty

[1] In development economics, the more general concept of 'informality' is used more often than 'undeclared', mainly because it encompasses a wider variety of instances, not only the one associated with tax evasion. However, the difference between the two is generally very subtle. In this paper we use both terms interchangeably.

[2] The debate in the literature revolves around the question whether informal and formal labour markets are segmented or integrated. Although this paper does not dig into this matter, it sympathises with the view of integrated labour markets and provides encouraging evidence in this respect.

[3] See also Lehmann & Tatsiramos (2012).

[4] The same approach has been used in Cappariello & Zizza (2010) and Capasso & Jappelli (2013)

laws targeting undeclared work, which was issued by the Italian government in 2001-02.[5] In particular, we use the exogenous variation in the supply of undeclared labour generated by the policies in order to build a valid instrument for the identification of the structural parameter of labour demand. The estimated elasticities are then used to describe wage inequality responses (the ratio of skilled to unskilled wages for declared workers) to fluctuations in the supply of undeclared work. Our main findings show that a rise in the supply of undeclared work (i) decreases undeclared wages, (ii) increases declared wages and (iii) reduces declared wage inequality. We find q-complementarity between undeclared workers and low to medium-skilled jobs.

The paper proceeds as follows: Section 2 describes the institutional setting in which the analysis is framed and gives a detailed account of the three tax amnesty laws. Section 3 illustrates the theoretical framework and Section 4 illustrates the empirical analysis and our identification strategy. Section 5 shows the results and a number of robustness checks. Section 6 concludes and indicates the main policy implications of our findings.

2. The Tax Amnesty Laws

Figure 1 depicts the evolution of log full time equivalent (FTE) undeclared and declared workers from 1990 to 2008 in Italy. The data is from the Italian Institute of Statistics (ISTAT), which regularly provides estimates of the Italian black market GDP and employment disaggregated by region and industry sector. At least two facts have emerged from Figure 1. Firstly, undeclared and declared employment are positively correlated and share a similar increasing trend. Secondly, the undeclared FTE series shows a remarkable drop of about 3% between 2001-2003 and a level shift from 2003 onwards. Note that, despite this fall-off, there is no considerable increase in the formal part of employment from 2001 to 2003. The drop in undeclared employment is the result of three different tax amnesty laws enacted by the Italian government in 2001 and 2002. The tax pardons are Law no. 383/2001, Article 33 of Law no. 189/2002 and Law no. 222/2002.[6]

- Figure 1 about here -

The aim of these laws was to offer an amnesty for evasion of social security payments as well as to regularise the status of undocumented immigrants. More specifically, the aim of Law 383/2001 was to foster the formalisation of undeclared workers irrespective of the their citizenship status, while Article 33 of Law 189/2002 and Law 222/2002 exclusively targeted immigrants.

[5] Gobbi & Zizza (2012) made use of the same shock induced by one of the policies considered in this paper (Law no. 189/2002) to evaluate whether the development of a local credit market is affected by the informal sector.

[6] Law no. 189/2002 (the so-called 'Bossi-Fini' law) is a large legislative package which establishes new rules regarding immigration and asylum issues.

Act 383/2001 provided a three-year fiscal and social security payment reduction to those employers who declared informal labour positions. In particular, only informal dependent job positions that had started before the 25th of October 2001, and still ongoing at the date of application, were allowed to apply. The regularised position could be on a temporary as well as permanent basis, and involving full- or part-time work schedules.

The last two acts were mainly concerned with preventing illegal immigration. These policies targeted non-native workers employed informally. Job positions that were deemed informal at least three months prior to the law coming into force, were eligible for the pardon. Employers for the new regularised jobs were obliged to give their employees at least a one year contract and they were also required to a pay 700 Euro fine for each declared job position. The only difference between the two tax amnesty laws concerned the type of work eligible for regularisation: Law 189/2002 targeted housemaids and healthcare workers, while Law 222/2002 dealt with all kinds of subordinate labour. An additional and very important feature of Laws 189/2002 and 222/2002 was that regularisation of the labour status also granted residence permits for immigrants. The government promoted the campaign of regularisation significantly, via the media and social security auditors, to inform all interested parties.

The effects of these laws on the employment dynamics are well described in Figure 1: the steadiness of the ratio of undeclared to declared FTE workers is temporary interrupted by a decrease in informal labour between 2001 and 2003, while formal employment seems not to deviate from its increasing trajectory. As the policies have an impact on the legal status of undeclared workers, we would expect a perfect reallocation from undeclared to declared employment. However, careful inspection of Figure 1 puts forth a different pattern: the drop in undeclared work is not as equal as the increase in declared work. Why is this so? Several factors can account for this finding. Firstly, it can be the result of a partial regularisation of the actual hours of work. Because of the desire to have a stable path of consumption, workers may have chosen to declare only part of the hours of work while leaving the rest still 'off the books', so as to match their previous level of income. Analogously, the employers may have favoured this option because it gave them the chance to abide by the tax laws while still benefiting from the underreporting of hours which were actually worked. Moreover, the lack of any enforcement mechanisms could also have contributed to this partial regularisation.

Secondly, it might be that most of the regularised jobs were short-lived and they were possibly not registered between the years which were surveyed. In fact, as reported in Anastasia *et al.* (2007), about 70% of regularised positions were destroyed before 2004. This is not surprising, as the main policies' goal was to increase government revenues by recovering evaded taxes and to tackle rising illegal immigration.

This context and the shock of informal labour which stemmed from the tax amnesties, enables us to identify labour demand elasticities for undeclared and declared work.

3. Theoretical Settings

In this section we briefly sketch the theoretical implications of a surge in the supply of undeclared work. The theoretical apparatus is a model of the demand side of the labour market.[7]. The aggregate production function is the well-known and popular Cobb-Douglas aggregation:

$$Y = AK^{\alpha}N^{1-\alpha} \tag{1}$$

where A is exogenous total factor productivity, K is the physical capital and N is a CES aggregate of two different types of labour, undeclared and declared. The labour aggregate is defined as:

$$N = [\theta_U N_U^{\rho} + \theta_D N_D^{\rho}]^{1/\rho} \tag{2}$$

where ρ is a function of the elasticity of substitution σ_{DU} between the two types of labour ($\rho = 1 - 1/\sigma_{DU}$), θ_U and θ_D are the share parameters summing to 1. The competitive market imposes that all factors are paid their marginal product, then the undeclared and declared wages are given by

$$\log w_U = \log\left[A\left(\frac{K}{N}\right)^{\alpha}(1-\alpha)\right] + \log\theta_U + \frac{1}{\sigma_{DU}-1}\log\left[\theta_U + \theta_D\left(\frac{N_D}{N_U}\right)^{\rho}\right] \tag{3}$$

and

$$\log w_D = \log\left[A\left(\frac{K}{N}\right)^{\alpha}(1-\alpha)\right] + \log\theta_D + \frac{1}{\sigma_{DU}-1}\log\left[\theta_U\left(\frac{N_D}{N_U}\right)^{-\rho} + \theta_D\right] \tag{4}$$

Given these equations, it is then straightforward to show the effects of an increase in undeclared employment over declared and undeclared wages. Taking the partial derivative of (3) and (4), we obtain the wage effect of an increase in undeclared employment. The resulting expressions are as follows:

$$\frac{\partial \log w_U}{\partial \log N_U} \equiv \frac{1}{\sigma_U} = -\alpha S_U - \frac{1}{\sigma_{DU}}(1 - S_U) \tag{5}$$

$$\frac{\partial \log w_D}{\partial \log N_U} \equiv \frac{1}{\sigma_D} = -\alpha S_U + \frac{1}{\sigma_{DU}} S_U \tag{6}$$

where $S_U = w_U N_U/(w_U N_U + w_D N_D)$ is the share of overall wages paid to undeclared workers. To discuss these expressions we need to take into account whether the supply of physical capital is fixed or can change elastically perfectly.

[7] This approach is standard in studies of wage inequality. See (Katz & Murphy, 1992; Katz & David, 1999; Goldin & Katz, 2009)

Let us first examine the case in which capital is fixed. An increase in the supply of undeclared workers reduces both declared and undeclared wages by lowering the capital-labour ratio of the economy. Furthermore, if undeclared and declared labour are perfect substitutes ($\sigma_{DU} \rightarrow \infty$), declared and undeclared wages decrease by the same amount αS_u. If, instead, undeclared and declared labour are imperfect substitutes (or q-complements), there is also a positive effect on declared wages operating by the term $\frac{1}{\sigma_{DU}} S_u$. Which of the two effects on declared wages prevails is an empirical matter.

Now considering the situation where capital is supplied elastically perfectly, so that $\alpha \rightarrow 0$, since capital can adjust freely to changes in labour, informal labour does not affect the capital-labour ratio and if the two labour inputs are perfect substitutes, no changes in wages occur. However, if we consider imperfect substitutability, declared wages increase and undeclared wages decrease.

The previous production function considers declared and undeclared labour as homogeneous input. We can relax this assumption by taking into account different broad education groups of the workforce. In doing so, we allow for labour input to be comprised of high-skilled declared workers and the aggregate between undeclared and low-skilled declared workers. The choice for considering undeclared labour as an aggregate is merely practical, because in the empirical investigation we have exploited a source of exogenous variation only in total full-time equivalent undeclared workers. The aggregate production function is still described by (1), while N is a CES aggregate of high-skilled declared labour and the composite input and it is defined as:

$$N = [\theta_H N_H^{\gamma} + \theta_A N_A^{\gamma}]^{1/\gamma} \tag{7}$$

where γ is a function of the elasticity of substitution σ_{HA} between the two types of labour ($\gamma = 1 - 1/\sigma_{HA}$), θ_H and θ_A are the share parameters summing to 1. Furthermore, we assume that the labour composite N_A is itself the CES sub-aggregate of undeclared labour and low-skilled declared labour. It is defined as:

$$N_A = [\theta_U N_U^{\eta} + \theta_D N_{LD}^{\eta}]^{1/\eta} \tag{8}$$

where η is a function of the elasticity of substitution σ_{U-LD}between undeclared and declared low-skilled workers. θ_U and θ_D are the corresponding relative efficiency parameters.

In a competitive market, the marginal product for each labour supply equates to the corresponding wage. Thus the ratio of the wage rate of high-skilled declared workers to the wage of low-skilled declared workers equates to the ratio of the corresponding marginal products, satisfying the following equation:

$$\log\left(\frac{w_H}{w_{LD}}\right) = \log\frac{\theta_H}{\theta_{LD}} + (\gamma - 1)\log N_H - (\gamma - \eta)\log N_A - \log\theta_D - (\eta - 1)\log N_{LD} \tag{9}$$

Differentiating equation (9) with respect to N_U, we obtain the effect of an increase in undeclared employment over declared wage inequality. The resulting expression is

$$\frac{\partial \log(w_H/w_{LD})}{\partial \log N_U} = (\eta - \gamma)\frac{S_U}{1 - S_H} \tag{10}$$

or equivalently

$$\frac{\partial \log(w_H/w_{LD})}{\partial \log N_U} = \left(\frac{1}{\sigma_{HA}} - \frac{1}{\sigma_{U-LD}}\right)\frac{S_U}{1 - S_H} \tag{11}$$

where S_U is the share of overall wages paid to undeclared workers and S_H is the share of overall wages paid to the high-skill declared workers. An increase in undeclared labour increases the declared wage skill premium if undeclared workers compete more with low-skilled declared than high-skilled declared workers, that is, when $\eta > \gamma$ or, equivalently, when the elasticity of substitution between undeclared and declared low-skilled workers is higher than the elasticity of substitution between high-skilled declared workers and the aggregate labour input.

4. Estimation Strategy

4.1. Data and Descriptives

The main results are obtained using a sample of individuals drawn from the, publicly available, Bank of Italy's Survey on Household Income and Wealth (SHIW). The temporary nature of the shift in undeclared work urges us to base the analysis on wave 2000 and 2004, that is in the period just before and just after the implementation of the tax amnesty laws.[8] The dataset provides information about individual characteristics such as hourly net wages, payment of social security contributions, education, years of potential work experience as well as socio-demographic traits. From this dataset, to detect workers employed in the informal sector we couple two indicators which are individual wages and payment of social security contributions. The latter information is taken from the following question: *Over the whole working lifetime did you or your employer pay any social security contributions, even for a short period of time?* People that report total labour earnings and answer negatively to this question are categorised as undeclared workers.[9] In addition, for those people answering affirmatively, SHIW reports the cumulative number of years in which the employer or the employee have paid social security contributions. In constructing

[8] We discard the possibility to use wave 2002 in place of 2004, because in such year the effects of the tax amnesty laws had not been realised yet.

[9] Total labour earnings do not include unemployment benefits.

the measure of undeclared work, we also exploit such information, that is, we include individuals whose the number of years covered by social security contributions amounts to less than one tenth of the number of years of potential work experience. By using this threshold, we are confident of singling out individuals who have been employed in the undeclared sector for most of their working life.[10] The key findings of the paper are based on this indicator. As a robustness check, we also estimate models where the indicator of informality does not include this last group of informal workers. This indicator can be deemed a more conservative measure of undeclared work.

We restricted the sample to men, aged 15-64, who had had paid employment within the private sector. The focus on men is meant to abstract from relative trends in female wages that are driven by changes in relative selectivity of female workers. While the attention paid to the private sector is triggered by the fact that the FTE undeclared and declared employment is only available for the private sector. However, to validate the robustness of our results we estimate models that also include women and the public sector.

We exclude self-employed people, since they tend to differ from employees in identifiable ways and may therefore yield a different elasticity of demand for labour. Moreover, the self-employed are able to tailor their hours of concealed work to their personal preferences. In contrast, employees are more constrained by employer's needs. Lastly, the tax amnesty laws were designed to disclose only employee jobs.[11]

Information on undeclared and declared employment are taken from ISTAT. In particular, we use the regional time series of declared and undeclared full-time equivalent workers for the time period 1995-2005 to construct a regional indicator of the relative supply of undeclared employment. We decided to use ISTAT's regional indicators instead of calculating our own from SHIW data, mainly because they cover a longer time frame. Unfortunately, the Bank of Italy's dataset is released every two years and information on the payment of social security contributions has only been included since 1995, leaving us with three data points from 1995, 1998, 2000. This is an insufficient number of observations for our estimation strategy, which hinges on the estimation of separate autoregressive models for relative supply of informal labour for each region (see section 4.3). Figure 2 plots the 1995-2005 regional averages of the ISTAT informality rate against the same index as computed from SHIW. It is apparent that the two measures are strongly correlated with a coefficient of 0.82, equivalent to the one found in Capasso & Jappelli (2013) and similar to that of Cappariello & Zizza (2010)'s. Both indicators reveal that the bulk of undeclared work is concentrated in southern regions, where Campania, Calabria, Sicilia, Sardegna and Puglia have the highest level (above 20%), while northern

[10] We have also experimented with different thresholds such as the 5th and the 3rd percentile of the distribution of the ratio between the number of years of social security contributions and the number of years in employment, but the main conclusions of the paper are unaffected.

[11] Notice that over the period 2000-2005, irregular self-employment accounts for about a fifth of total undeclared labour (ISTAT, 2008).

regions employ a smaller share of informal work.

An important issue with ISTAT's regional time series is that it does not provide separate measures for dependent employment and self-employment. This poses some difficulties for the present analysis since our interest is to estimate wage elasticities of labour demand for employees. If, for example, changes in the level of irregular employment are also due to shifts in irregular self-employment, then using the indicator from ISTAT may overstate the effects of the tax pardons and lead to bias estimates of the wage elasticity. To eliminate such a possibility, we compare the informality rate from ISTAT and an indicator of irregularity for undeclared employees from SHIW (Figure 3). A visual inspection of the scatterplot suggests that the distribution of irregular self-employment follows very closely that of irregular dependent employment. As a consequence, the variation across regions of the dependent employment is almost entirely explained by the variation in the total irregular employment. Since there are cases where perfect one-to-one relationships can be rejected, notably Calabria and Emilia-Romagna, we include region and year fixed effects in all models. To put it differently, as long as both regular and irregular self-employment are constant over time, i.e. are not affected by the tax amnesty laws, they can be interpreted as measurement errors in the value of declared and undeclared dependent labour and can be ironed out by means of regional fixed effects. In the current analysis, the assumption of stable dynamics for regular and irregular self-employment is not as severe as it might seem, given that all models are estimated over a limited time frame, i.e. 2000-2004, which is insufficient to witness large shifts in the variables. In fact (ISTAT, 2008) reports that the irregularity rate for self-employment even slightly increased, going from 8.5% in 2001 to 8.7% in 2005. Therefore, the use of the ISTAT measure of irregular overall employment does not seems to be an important concern.

In this paper the wage outcomes are constructed by dividing annual wage and salary earnings by the product of weeks worked and usual hours per week. Earnings in SHIW are reported net of both labour and income taxes, and unfortunately there is no possibility to infer the gross analogue. Of course, only wages for declared employees are actually netted. The use of after-tax wages poses some concerns for the estimation of the own-wage and cross-wage elasticities that need to be duly addressed.

Social security contributions are determined as a percentage of taxable income and the contribution rate can vary mainly according to the sector of activity, type of occupation (e.g., blue-collar, white-collar, manager) and firm size (the categories are 1-15, 15-50 and 50+ employees). This can be regarded as classical measurement error in the net earnings variable, i.e. error is uncorrelated with gross earnings and other explanatory variables, leading to less precise but still consistent estimates.

A different difficulty is related to income tax. The Italian tax system is progressive and during the period 2000-2004 it had 5 marginal tax rates ranging from 18% to 45%. The system also allowed for a number of deductions and allowances which basically related to the taxpayer's family status (e.g., whether she/he had dependent children, a partner or other dependent relatives), health care, education and home ownership. So what would be the appropriate tax rate for use in the calculation of final taxation, hinges on a complex model and on gross earnings. To add further complications, during that period under scrutiny, the marginal tax rates also underwent several reforms. As a result, measurement error is likely to be correlated with both gross earnings and household characteristics, and to lead to biased estimates. Bound *et al.* (1994) have found that this bias is generally small and it vanishes as the variance of the true earnings increases. Also, Bound & Krueger (1991) came to a similar conclusion that mismeasurement of earnings gives rise to little bias when earnings are used as a left-hand side variable of a regression. Moreover, the instrumental variables strategy adopted in this paper to cure simultaneity bias, is most likely to be helpful for correcting the bias arising from measurement error also.

However, to dismiss the possibility that our results are driven by the use of net earnings, we estimate models based on imputed gross wages. We assign before-tax wages to each declared worker in the SHIW sample through the predictive mean matching procedure. Gross wages are gleaned from European Community Household Panel data (ECHP) for the year 2000, and the Italian version of the European Union Statistics on Income and Living Conditions dataset (IT-SILC) for 2004.[12]

Imputation by predictive mean matching borrows an observed value from a donor with a similar predictive mean. Specifically the procedure works as follows: first, a linear regression model is estimated to obtain linear predictions; the linear predictions are then used as a distance measure to form the set of nearest neighbours consisting of complete values; finally it randomly draws an imputed value from this set.[13] The procedure is then replicated 100 times. For each of the 100 datasets we estimate our empirical models and calculate the coefficients' (and standard errors') means. Notice that the variance estimate of the coefficient means will reflect not only sampling variability but also our uncertainty about the imputation model.[14] The linear regression model controls more than 40 variables, and most of them are strong predictors of gross income. These are net wages, age, gender, marital status, a dummy for dependent children, education, potential work experience and its square, interactions between experience and education levels, sectors of activity, a dummy for part-time work schedule, a dummy for fixed-term job, home ownership, a dummy for migrants and region of residence. Net wage is a strong predictor for gross wages and its

[12]In fact, we use wave 2000 of ECHP that reports information on current gross income and wave 2005 for IT-SILC that has information of previous year income.

[13]We have experimented with different set of possible donors, i.e. considering 1, 3 and 5 values, but we have not detected any substantial differences.

[14]The estimator of the variance V_β is given by:

$$V_\beta = \frac{1}{m}\sum_{m=1}^{M} s_m^2 + \Big(1 + \frac{1}{m}\Big)B$$

where $B = \frac{1}{m-1}\sum_m (\hat{\beta}_m - \hat{\beta})^2$ so as to reflect variation within and between imputation.

inclusion in the regression model should allay concerns for misspecification due to omitting relevant variables.

Another possible issue with earnings data is misreporting. In particular, for fear of being detected by the tax authorities, undeclared workers may be tilted to declare that they do not earn any income when they, in fact, do. This would cause our approach to detect only part of people working in the underground economy and our results to be based on a restricted sample of undeclared workers. While we cannot exclude this possibility, we are confident that such issue is fairly mitigated by the fact that (i) the survey is anonymous and (ii) if caught by the tax authorities, undeclared dependent workers are not prosecuted for working irregularly. Only the employers would end up in court. To further alleviate the issue of misreporting, we have carried out an additional check, based on the subjective assessment of the interviewer about the trustworthiness of income and wealth information. Specifically interviewers are asked to give a score, on a 1 to 10 scale, where 10 is best, responding to the question: *To what extent do you think information given by the interviewee on income and wealth are true?*. We have found that the average score for the group of declared workers is equal to 7.7 and, most importantly, is not statistically different from the average score of 7.5 for undeclared workers.

Our final sample consists of $11,965$ individuals, evenly distributed across the two waves. Table 1 contains descriptive statistics for the full sample (including men, women, workers of both private and public sector) and reports figures for declared and undeclared workers separately. Informality is substantial for prime-age workers (50% is concentrated in the 25-39 age group), men (59%) and unmarried people (51%). Interestingly, while immigrants cover a little share of declared jobs (4%), their participation in the irregular economy is about 12% against 88% of native citizens. Participation rate in the irregular economy is a decreasing function of the level of education attained; 54% of undeclared workers have compulsory education (less than High School degree), about 35% have a High School diploma and only 11% have a College degree. Workers in the regular economy show on average a high labour market experience, 60% have between 11 and 30 years of work experience; while people working irregularly seem to have less work experience, 60% of them are found in the group 1-20. According to our sample, the bulk of irregular economy is concentrated in the Manufacturing (21%), Trade (18%) and Services sector (15%). A remarkable part of undeclared work (19%) is also present in the Public administration, Defence, Education and Health care sector. Then, the effect of firm size on the participation rate in the undeclared labour market is decreasing with the number of employees; more than a half of the undeclared population are employed by small-sized enterprises (who employ between 1 and 19 employees). In accordance with the official ISTAT's estimates of the underground economy, our sample indicates that southern regions report the highest level of undeclared employment (45%), followed by the northern regions (33%) and the centre of Italy (22%)

The bottom of Table 1 reports information on net hourly wage for both undeclared and declared workers. As we can see, the proportion of employees in the declared economy across income quartiles is virtually constant, possibly reflecting the progressivity of the tax system; while mean hourly wage is about 2 log points with a standard deviation of 0.4. In the irregular sector people seem to be more concentrated in the lower end of the wage distribution; in fact, more than two third of undeclared workers earns an income below the median value, and hourly wage amounts on average to 1.8 log points with a standard deviation of 0.52.

4.2. Empirical Models

To obtain the elasticities of demand and substitution as envisaged by the theory, we estimate models of the form:

$$\log w_{irt} = \alpha_r + \zeta_{2004} + b_i + X'_{irt}\beta_t^b + \gamma \log\left(\frac{U_{rt}}{D_{rt}}\right) + \delta b_i \log\left(\frac{U_{rt}}{D_{rt}}\right) + \varepsilon_{irt} \quad (12)$$

where the left-hand-side variable, $\log w_{irt}$, is log hourly net wage of workers in 2000 and 2004, two years before and after the tax amnesty laws were implemented. On the right-hand side, α_r are regional fixed effects, ζ_{2004} is a time dummy, b_i is a dummy for undeclared status, and X'_{irt} is a vector of covariates that controls for individual characteristics. These are potential work experience and its square, education level, interactions between experience and education level, a dummy for migrants and a dummy for living in urban areas. Each of the individual controls are permitted to affect declared and undeclared worker's earnings differentially by undeclared status and year. γ and δ are the coefficients of major interest for our research. γ is the coefficient of $\log \frac{U_{rt}}{D_{rt}}$ which is the logarithm of the ratio of regional FTE undeclared workers to regional FTE declared workers, and δ is the coefficient of the interaction term between b_i and $\log \frac{U_{rt}}{D_{rt}}$.[15]

Of course, the ratio $\log \frac{U_{rt}}{D_{rt}}$ represents equilibrium relative supply of undeclared work. If not duly tackled, the coefficients γ and δ will be affected by simultaneity bias. As we are interested in estimating labour demand functions, we need an exogenous variation in relative supply of undeclared work. Section 4.3 shows our identification strategy and explains how the tax amnesty laws are used to construct a reliable instrument. Once the identification problem is solved, we can interpret the coefficients γ and δ as follows: γ is the effect of a shift in relative supply of undeclared labour on both declared and undeclared earnings and δ measures the differential effect of informal labour supply on informal wages; therefore, δ is an estimate of the inverse elasticity of substitution between the two types of labour. The sum of γ and δ describes how a change in the relative supply of undeclared workers affects undeclared workers' wages and represents an estimate of the inverse of own-wage elasticity for undeclared labour. To account for the fact that relative supply of undeclared

[15] A similar approach can be found in Acemoglu *et al.* (2004).

labour has group-wise structure, we report robust standard errors clustered on a region-year level.

The impact of undeclared labour supply may not be uniform throughout the distribution of declared wages. For instance, wage effects of an increase in informality are expected to be more negative for unskilled declared workers than for skilled regular workers if undeclared labour is a closer substitute for unskilled declared labour. This could also have consequences on the dynamics of declared earnings inequality (the ratio of skilled to unskilled wage). To investigate this possibility we consider the following equation:

$$\begin{aligned}\log w_{irt}^{D} \quad = \quad & \alpha_r + \zeta_{2004} + h_i + X'_{irt}\beta_t^h + \pi \log\left(\frac{U_{rt}}{D_{rt}}\right) + \lambda h_i \log\left(\frac{U_{rt}}{D_{rt}}\right) + \\ & \theta \log\left(\frac{H_{rt}}{L_{rt}}\right) + \lambda h_i \log\left(\frac{H_{rt}}{L_{rt}}\right) + \varepsilon_{irt} \qquad (13)\end{aligned}$$

Model (13) is very similar to model (12) but here we have only used data for declared people. Therefore, $\log w_{irt}^{D}$ is log hourly net wage of declared workers in 2000 and 2004. We still control for regional and time fixed effects and the vector X'_{irt} of individuals controls. All variables are allowed to have different effects on the earnings of high- and low-skilled declared workers and to differ by year. h_i represents a dummy for the high-skilled group of workers and $\log \frac{H_{rt}}{L_{rt}}$ is the log relative supply of high-skilled versus low-skilled declared labour. We use again robust standard errors clustered by region and year of observation.

The coefficients of interest in this equation are π and λ. Following the same reasoning as in the previous model, and once the identification problem has been dealt with, π can be interpreted as the effect of a shift in the relative supply on the wages of low-skilled declared workers. Keeping the employment levels of high- and low-skilled declared workers constant, π is also an estimate of the inverse of the cross-wage elasticity of demand between undeclared labour and low-skilled declared labour. Similarly, the sum of π and λ measures the impact of undeclared labour supply on the earnings of high-skilled declared workers, and the inverse of it is an estimate of the cross-elasticity of demand between undeclared and high-skilled labour. Lastly, the ratio between the two cross-wage elasticities, $(\pi + \lambda)/\pi$, tells us which education group, undeclared labour is a closer substitute for. Specifically, if $\pi + \lambda/\pi < 1$, it implies that undeclared labour has a larger wage impact on the least skilled group; therefore undeclared workers are closer substitutes for the least skilled people than for the highest skilled group of declared workers. The opposite holds if $\pi + \lambda/\pi > 1$.

4.3. Identification

We have identified the labour demand function by using the tax amnesty laws brought into force in 2001/2002. As Figure 1 has shown, these measures produced an aggregate labour supply shock of undeclared work. The reduction of undeclared labour did not lead to the creation of stable employment; what we observe from annual data is that most of the regularised jobs were destroyed shortly after the regularisation. We take advantage of this natural experiment to study the short-run adjustment of the labour market.

Identification of the labour demand elasticities requires an instrumental variable, $Z_{r,t}$, that is associated with changes in the relative supply of undeclared work but do not lead to change in the dependent variable, $\log w_{irt}$. Proven that $Z_{r,t}$ is an instrument for $\log \frac{U_{rt}}{D_{rt}}$, it follows that $b_i \times Z_{r,t}$ is a good instrument for $b_i \times \frac{U_{rt}}{D_{rt}}$.

We consider an instrument of this form:

$$Z_{r,t} = \zeta_{2004} \times \log A_{r,2002} \tag{14}$$

where $\log A_{r,2002} = \log(\widehat{FTE}_{und} - FTE_{und})_{r,2002}$. $\widehat{FTE}_{und}$ is the prediction of FTE workers in the irregular economy in 2002, i.e. regional undeclared employment in the absence of policy interventions. From this quantity we subtracted the actual values for informal FTE workers, multiplied it by a time dummy for 2004 and then took logs. Therefore $\log A_{r,2002}$ is a proxy for the number of regularization.[16]

To predict the counterfactual value for undeclared FTE employment, we use an autoregressive model of order one, AR(1), one for each regions.[17] Table 2 summarises the estimates of these models. Columns 2-3 report coefficients and standard errors, columns 3-4 depict goodness-of-fit measures (R-squared and root-mean-squared-error) to assess the performance of the prediction models. All coefficients are less than 1 in absolute value, suggesting that the time series are stationary, and most of them are significantly different from zero at conventional levels. The AR(1) models also appear to fit the data well. According to the R-squared, the best fit is obtained for southern and central regions, where the presence of irregular work is higher, while in terms of root-mean-square-error, all models give quite an accurate prediction of undeclared employment.

Before turning to the discussion of the results, we need to address a possible question: to what extent is the instrument capturing a supply-side shock? If, for instance, changes in the level of undeclared employment are due to changes in the demand for informal work, then our approach will fail in the attempt to identify the wage elasticities for labour demand. The prospect of legal status constituted a strong incentive for employees to ask for the regularisation of their job positions. This incentive was stronger for immigrants, who had even more to gain from the regularisation, i.e. acquisition of a residency permit, full entitlement to welfare benefits and access to the national health service. On the employer's side, the will to declare concealed job positions may have

[16]Data on actual regularisations are sparsely available and we are not aware of any public source where information on all programmes considered in this paper are reported.

[17]In a similar vein, Card (2009) uses predicted inflows of immigrants to instrument the relative supply of migrants versus native workers in the USA. Pesaran & Smith (2012) estimate an ARDL model to construct counterfactual scenario and evaluate the effect of Bank of England's quantitative easing on UK output growth.

been responding to the need to show compliance with tax regulations in the expectation of reducing the chance of being audited by tax authorities, while still having room to keep workers 'off the books'

- Table 2 about here

5. Results

In this section we report our estimation results for various versions of the models outlined above. Section 5.1 shows the main estimates of the own and cross-wage elasticity of labour demand; section 5.2 describes our findings concerning the impact of undeclared work on declared earnings inequality; section 5.3 presents estimates of equation (12) and (13) when imputed gross earnings for regular workers is used in place of after-tax earnings. Finally, section 5.4 provides some robustness checks.

5.1. Own- and Cross-wage Elasticity of Labour Demand

We start with the ordinary least squares (OLS) estimates of equation 12. The point estimates in the first two rows of Table 3 correspond to γ and δ. Column I is the most parsimonious model that only controls for the relative supply of undeclared workers, regional and time dummies. Models in columns II and III add a set of individual characteristics and interactions with the undeclared worker dummy to the baseline specification. Column III also includes interactions with a 2004 dummy. Column IV controls for regional covariates such as the share of undeclared migrants, the percentage of undeclared young workers, average education, the share of undeclared workers in agriculture, construction and manufacturing, all measured in 2000 and interacted with the dummy for 2004. These regional variables are meant to control for the fact that reduction in undeclared labour supply may have, in part, been triggered by sectoral distribution and the irregular workforce's demographic structure, which, presumably, are correlated with their analogues of the declared workforce. Finally, in column IV we have augmented the model with the lag of regional average undeclared earnings and the percentage of undeclared earnings found in the public sector in 2000.[18]

The estimates for γ and δ in Table 3 reveal the negative impact of a shift in relative supply of undeclared work on undeclared wages. According to columns II-III a 1% increase in relative supply decreases wages for irregular workers by 0.06-0.1%. When we include regional controls (column IV-V) the OLS estimates indicate that a 1% increase in relative undeclared employment raises undeclared wages by 0.07%. However, despite being informative, findings in Table 3 are

[18]It might be surprising that public sector could employ people irregularly, however it is not uncommon to find politicians' personal assistants and collaborators of managers of public companies who work without a formal labour contract. Therefore, our variable most likely captures those individuals or others found in similar circumstances.

affected by simultaneity bias and they are most likely to be biased estimates of the true parameters.

- Table 3 about here -

To obtain unbiased estimates, we apply two-stage least squares for the estimation of equation 12. Our empirical results for the undeclared and declared male wages are presented in Table 4. The columns differ in that we progressively add controls and are arranged as in Table 3. Column I does not use any controls, except for region and time dummies. Column II includes individual controls and interactions with undeclared status. Column III adds in interactions with a 2004 dummy. Column IV brings in the first set of regional controls: the share of undeclared migrants, the percentage of undeclared young workers, average education, the share of undeclared workers in agriculture, construction and manufacturing, all of which were measured in 2000 and interacted with the dummy for 2004. Columns V factors in other regional characteristics such as past undeclared earnings at the regional level and the percentage of undeclared found in the public sector in 2000. To test whether the coefficients γ and δ are jointly statistically significant, we report in Table 4 p-value of the F statistic for joint significance.

To check the validity of the instruments, the bottom of Table 4 presents first-stage diagnostics. In particular, we show tests of underidentification and weak identification for each endogenous regressor using the method suggested by Angrist & Pischke (2008) (AP, henceforth). For the AP underidentification test we report p-values, while the F-stat for the AP weak identification test is contrasted with the Stock & Yogo (2005) critical value based on rejection rate of 10%. In all models, the null of underidentification is rejected at conventional level. The AP F-stat also indicates an absence of weak instruments: in fact, the value of the statistic is, in all cases but one, greater than the critical value of 16. Taken together, these tests reassure us about the validity of our instruments.

The first row of Table 4 shows the estimates of the coefficient associated with γ in equation 12. They are positive and significantly different from zero only in column IV and V, that is, when additional variation at the regional level have been taken into account. This suggests that the bulk of variation in relative supply variable is found at the regional level. Differences in sectoral distribution and demographic structure of the irregular workforce across regions prove to be important predictors of the shift in relative supply observed from 2000 to 2004. According to columns IV and V, a 1% increase in relative supply of informal work raises overall wages by 0.3%.

Summing δ and γ we obtain the effect of an increase in the relative supply of undeclared workers on undeclared wages. The wage effect is significantly different from zero in column II and III, while it is estimated less precisely as we brought in regional controls (column IV and V). A surge of 1 percentage point in relative supply decreases undeclared wages by 0.3-0.4% (column II and III), However, these wage effects are almost negligible in models in column IV and V. The point estimates of δ and γ imply an elasticity of demand for undeclared labor in the range of 2.5-2.7 in absolute value, with larger elasticity in the models which include regional covariates.

We obtain an estimate of the elasticity of substitution between undeclared and declared labour by looking at the coefficient in the second row of Table 4, corresponding to δ in equation 12. The estimates of δ are, in all cases, negative, significantly different from zero, and strongly stable across all specifications. We find that an increase of one percentage point in relative supply generates a drop in the wage differential between regular and irregular workers equalling 0.4%. This value corresponds with an estimate of 2.4 in the elasticity of substitution, σ_{DU} in absolute value. Taken together, these findings provide encouraging support for the idea that a reduction of the undeclared sector may decrease the level of wages of declared workers and probably the overall wage inequality, i.e., the one including undeclared and declared earnings.

- Table 4 about here -

5.2. The Effect of Undeclared Work on Declared Wage Inequality

The estimates in Table 4 show that changes in undeclared labour supply have a non-negligible impact on earnings of declared workers. However, this wage effect might not be identical for all individuals in the declared economy. For instance, if undeclared labour is closer substitute for unskilled declared labour, the wage effect of an increase in undeclared labour supply will be more negative to unskilled than skilled workers. This may, in turn, leads to higher wage inequality for the declared workforce. To explore this possibility, we estimate various version of equation (13) for different educational groups. We split the sample into three broad education groups: college graduates (CLG), high school (HS) and lower than high school (LHS). CLG includes individuals who have either a college or higher education degree. HS is people who have a high school diploma and LHS considers individuals who have less than a high school diploma and no education. We carried out separate analyses for HS-LHS and CLG-HS group.

Our findings are reported in Table 5 and 6. We have only reported estimated coefficients for models that control for individuals characteristics and interactions terms, and include regional controls, given that baseline models (the analogues of model I and II in Table 4) do not bring additional information. Columns I and IV include individual characteristics and interactions with a 2004 dummy and high school (college) dummy; columns II and V takes regional covariates into account (the share of undeclared migrants, the percentage of undeclared young workers, average education, the share of undeclared workers in agriculture, construction and manufacturing, all measured in 2000 and interacted with the dummy for 2004); columns III and VI also controls for the lag of regional undeclared earnings and the fraction of irregular workers found in the public sector. Relative supply of regular workers with a high school education diploma versus people with a level of education which is less than high school (HS/LHS) should directly affect the HS/LHS wage premium. Analogously, relative supply of college versus high school declared graduates should impact on the CLG/HS wage premium. Therefore these quantities must appear in the estimation of wage equation (13). Columns I to III of Table 5 control for relative HS/LHS supply, and columns I to III of Table 6 include relative supply of college versus high school graduates. However, we also consider the case wherein relative supply measures are part of the error term, and uncorrelated with the instrumented relative supply of undeclared labour. We present the results of this exercise in columns IV to VI of Tables 5 and 6.

We shall start with the instrumental variables regressions for the HS-LHS sample in Table 5. The validity of our instruments is borne out by the first-stage diagnostics: the AP χ^2 test always rejects the null of underidentification, while the F-stat indicates the presence of weak identification for relative supply interacted with a dummy for high school.

As we can see, the coefficients associated with relative supply of undeclared work are consistently positive and significant in most cases. This indicates positive wage effects on people with both low and high levels of education (the null hypothesis of joint insignificance of $\pi + \lambda$ is always rejected at conventional levels). However, these impacts are not uniform across the two educational groups. In fact, the wage effects are more pronounced for the HS graduates than for people with a lower level of education. More specifically, a 1% rise in undeclared relative supply increases LHS wages by 0.3-0.7% and HS wage by 0.6-1%. This implies that the HS-LHS wage premium also increases by an amount of 0.2-0.4%, even though this is imprecisely estimated. The exclusion of relative supply of high school graduates in columns IV to VI seems to alter slightly the point estimates, but does not change their economic interpretation.

The ratio of cross wage elasticities of undeclared labour for HS versus LHS workers, $\sigma_{u-hs}/\sigma_{u-lhs}$, is equal to 1.7-2.2, consistent with the aforementioned stronger wage impact high school graduates. Since these quantities are greater than 1, we conclude that informal work has a higher degree of complementarity with medium skilled labour than low skilled labour.

Table 6 displays analogous estimates of equation 13 for the college-high school wage premium. The wage effects of a rise in undeclared work on college graduates are given by the sum of the coefficients π and λ, presented in the first two rows of Table 6. The results are not significantly different from zero given that we never reject the null hypothesis of the joint significance test. This suggests that informal labour has no influence on the highly skilled part of the labour force. Columns II, III, V and VI still capture positive effects on high school graduates, although they are smaller than Table 5.

- Tables 5 and 6 about here -

5.3. Using Imputed Gross Earnings

The use of after-tax wages as a left-hand-side variable may create some concerns for the estimation of the own-wage and cross-wage elasticities. As discussed in section 4.1, this may lead to biased estimates. Although the instrumental variables strategy adopted throughout the paper most likely solves this, we test the robustness of the results by estimating models similar to (12) and (13) but replace the dependent variable with imputed gross wages.

Table 7 contains estimated labour demand for undeclared and declared people and Table 8 showcases separate models for the CLG-HS and the HS-LHS group. Coefficients and standard errors have means of 100 iterations of the same model, each time with different imputed values for gross wages. We cannot calculate first-stage diagnostics but we report p-values for the joint significance $\gamma + \delta$ and $\pi + \lambda$ following the procedure outlined in Appendix A.

As we can see, both the size and the significance of the coefficients are virtually unchanged. This test reinforces our confidence in the previous empirical results: a rise in relative supply of undeclared labour positively affects both HS and LHS wages but has no impact of CLG earnings.

5.4. Robustness Checks

In this section we carry out a number of robustness checks. First, we examined whether the results in Tables 4 to 6 remain stable after women and public sector employees have been included. Table 9 shows estimates of equation 12 and reports only models that control for individuals characteristics and interactions terms, and include regional controls, comparable to column IV and V in Table 4. The point estimates of relative supply have a smaller size with respect of Table 4 and are statistically significant at conventional levels. In particular, we obtain that a 1% rise in relative supply of undeclared work reduces undeclared earnings by 0.06-0.2%, although this is imprecisely estimated, and increases declared wages, on average, by 0.2%, other things being equal.

The results of the impact of undeclared work on different educational groups of declared individuals are presented in Table 10. Column I and III consider women and men employed in the private sector, while models II and IV bring in people working in the public sector. As it is apparent, we still identify education-specific effects of an increase in relative supply of informal work: a positive impact on both LHS and HS earnings and no effect on CLG graduates.

- Tables 9 and 10 about here -

A second concern arises from our measure of undeclared labour. As explained in section 4.1, our main indicator for undeclared labour consists of people who had not, for a significant fraction of their working career, paid any social security contributions. We also constructed a second indicator which keeps only cases of people who had never paid any contributions. In Tables 11 and 12, we replicate models in Tables 4 to 6 but replace the indicator of undeclared work with this more conservative measure. As we can see the estimated wage elasticities have the expected sign but are less precisely estimated. Applying the more conservative measure of undeclared work, we obtain slightly smaller coefficients for γ and δ. The same takes place with the estimation of π and λ in Table 12. This does not, however, undermine our conclusions of a positive effect of undeclared labour on declared wages.

- Tables 11 and 12 about here -

6. Conclusions and Discussion

Our paper provides estimates of demand for labour when undeclared work is factored in. We exploit a quasi-experimental setting created by by three tax amnesty laws in 2001 and 2002 to calculate own-wage and cross-wage elasticities of labour demand. Our results reveal that a growth of 1% in relative supply of undeclared work decreases undeclared earnings by a minimum of 0.03% to a maximum of 0.4%. In addition, it increases wages for declared workers with medium and low levels of education but does not affect earnings for high skilled individuals. This indicates that undeclared labour is complement to low and medium skilled inputs.

Various theoretical explanations may account for the strong patterns elicited in the analysis. Firstly, comparative advantage and tasks specialisation can explain both the presence of a non-negligible irregular sector and complementary between formal and informal labour. Undeclared workers might have a comparative advantage in occupations requiring manual labour-intensive tasks, while declared workers have an advantage in high skill intensive tasks. Furthermore, the marginal cost of some manual occupations turns out to be less than or equal to marginal productivity only when it hasn't been declared to the tax authorities. Therefore, growth in the irregular sector will result in increased demand for regular workers, and higher earnings for low and medium skilled people. This is closely related to D'Amuri & Peri (2011)'s evidence of a shift in the occupational distribution of European natives toward more complex jobs, in response to low-skill migration.

Secondly, undeclared labour can have an impact on declared workers' time-use decisions, especially for people with high opportunity cost of time. Considering that a significant fraction of undeclared employment is found in services that are close substitutes for household production, we might expect declared workers to change their labour supply decisions on the intensive margin, e.g. increasing the time devoted to their main job. This shift will eventually result in higher wages. A similar conclusion is borne out by the work of Cortes & Tessada (2011), who find that an increase in the number of less skilled immigrants have caused the labour supply of high skilled women in the US to soar.

Finally, the role of trade unions in setting wages may help explain the differential impact of irregular work on the declared workforce. Normally, a trade union represents the interests of manual workers, and they have substantial bargaining power for setting wages. The insider-outsider theory provides explanations of why firms are reluctant to replace high-wage unionised employees with low-wage nonunionised ones. These explanations can be easily extended to the formal versus informal sector employees divide.

An important supplementary result of our analysis is about the size and direction of wage inequality responses to shift in the supply of undeclared work. The fact that undeclared work shows a strong degree of complementarity to low and medium skilled labour suggests that a reduction of informality might yield an increase in wage inequality for the declared part of the workforce. Figure 4 provides a visual check for this statement. We display the evolution of the Gini index, a popular measure of inequality, calculated from SHIW over the period 1995 to 2006. As we can see, the overall decline in wage inequality is interrupted by a one-off sharp increase in the Gini index when the tax amnesty laws came into force in 2001/2002. What is more, given the differential wage effects for medium and low skilled earnings we should be able to detect divergent growth of upper-tail and lower-tail wage inequality. In Figure 5, we display the trajectories of the 90/50 and 50/10 log hourly wage differential along with changes in overall wage inequality, summarised by the 90/10 log hourly wage gaps. As it is standard in the literature, the 90th, 50th and 10th wage percentiles are proxys for earnings of low, medium and highly educated individuals. The 90/10 wage inequality mirrors the dynamics of the Gini index presented in Figure 4. The 50/10 ratio indicates a constant reduction in the wage gap between HS and LHS graduates, while the 90/50 index reflects a worsening of the upper-tail wage inequality from 2000 to 2002, during the application of the tax amnesties. These patterns are strikingly consistent with our findings, and provide further encouraging, empirical support to the idea that the declared and undeclared part of the labour market are more closely related than is normally realised.

Also, our analysis can contribute to current debate on who benefits from tax evasion. Along with the standard truism that wants the main beneficiaries of successful tax evasion be the tax evaders themselves, Alm & Sennoga (2010) put forward a different mechanism under which benefits from tax evasion in the form of welfare increases are distributed over a broader group of individuals. Results presented in this paper seem to complement such a proposition very well: we find that earnings of low- and medium-skilled workers will soar in the aftermath of an increase in undeclared labour. This provides a possible explanation of the paucity of policies aiming to narrow down informality. Such policies might be stymied by the declared workers' unwillingness to accept welfare losses.

References

Acemoglu, Daron, Autor, David, & Lyle, David. 2004. Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. *Journal of Political Economy*, **112**(3), 497–551.

Alm, James, & Sennoga, Edward B. 2010. Mobility, competition, and the distributional effects of tax evasion. *National Tax Journal*, **63**(4), 1055–1084.

Anastasia, Bruno, Gambuzza, Maurizio, & Rasera, Maurizio. 2007. Gli immigrati regolarizzati nel 2002 e la continuit del loro impiego. *Economia & Lavoro*, **2**(maggio-agosto), 161.

Angrist, Joshua D, & Pischke, Jörn-Steffen. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bound, John, & Krueger, Alan B. 1991. The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics*, **9**(1), 1–24.

Bound, John, Brown, Charles, Duncan, Greg J, & Rodgers, Willard L. 1994. Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data. *Journal of Labor Economics*, **12**(3), 345–68.

Capasso, Salvatore, & Jappelli, Tullio. 2013. Financial development and the underground economy. *Journal of Development Economics*, **101**, 167–178.

Cappariello, Rita, & Zizza, Roberta. 2010. Dropping the books and working off the books. *Labour*, **24**(2), 139–162.

Card, David. 2009. Immigration and Inequality. *The American Economic Review: Papers & Proceedings*, **99**(2), 1–21.

Cortes, Patricia, & Tessada, José. 2011. Low-skilled immigration and the labor supply of highly skilled women. *American Economic Journal: Applied Economics*, **3**(3), 88–123.

D'Amuri, Francesco, & Peri, Giovanni. 2011. Immigration, jobs and employment protection: Evidence from Europe. *NBER, Working Papers*, **17139**.

Gobbi, Giorgio, & Zizza, Roberta. 2012. Does the underground economy hold back financial deepening? Evidence from the Italian credit market. *Economia Marche / Journal of Applied Economics*, **XXXI**(1), 1–29.

Goldin, Claudia Dale, & Katz, Lawrence F. 2009. *The race between education and technology*. Harvard University Press.

ISTAT. 2008. *La misura dell?occupazione non regolare nelle stime di contabilit nazionale*. Tech. rept. ISTAT.

Jütting, Johannes P, & Laiglesia, Juan R. 2009. Employment, poverty reduction and development: What?s new. *Is informal normal*, 17–26.

Katz, Lawrence F, & David, A. 1999. Changes in the wage structure and earnings inequality. *Handbook of labor economics*, **3**, 1463–1555.

Katz, Lawrence F, & Murphy, Kevin M. 1992. Changes in relative wages, 1963-1987: Supply and demand factors. *Quarterly Journal of Economics*, **107**(1), 35–78.

Lehmann, Hartmut, & Tatsiramos, Konstantinos. 2012. *Informal Employment in Emerging and Transition Economies*. Vol. 34. Emerald Group Publishing.

Pesaran, M Hashem, & Smith, Ron P. 2012. Counterfactual analysis in macroeconometrics: An empirical investigation into the effects of quantitative easing. *IZA Discussion Paper Series*, **6618**.

Schneider, Friedrich, & Enste, Dominik H. 2000. Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature*, **38**(1), 77–114.

Stock, James, & Yogo, Motohiro. 2005. Testing for Weak Instruments in Linear IV Regression. *Pages 80–108 of:* Andrews, Donald W.K. (ed), *Identification and Inference for Econometric Models*. New York: Cambridge University Press.

Appendix A. Reparametrization of models with imputed gross wages for declared workers.

To test the joint significance of $\gamma + \delta$, we reparametrize empirical models 12 and 13. For brevity's sake, this appendix only reports the reparametrization of model 12. Of course, the same calculation applies to equation 13.

Adding and subtracting $\gamma b_i \ln \left(\frac{U_{rt}}{D_{rt}} \right)$ from equation 12 and rearranging, we obtain:

$$\begin{aligned} \ln w_{irt} &= \alpha_r + \zeta_{2004} + b_i + X'_{irt}\beta_t^b + \gamma \ln \left(\frac{U_{rt}}{D_{rt}} \right) + \delta b_i \ln \left(\frac{U_{rt}}{D_{rt}} \right) \\ &+ \gamma b_i \ln \left(\frac{U_{rt}}{D_{rt}} \right) - \gamma b_i \ln \left(\frac{U_{rt}}{D_{rt}} \right) + \varepsilon_{irt} \\ \ln w_{irt} &= \alpha_r + \zeta_{2004} + b_i + X'_{irt}\beta_t^b + \gamma \ln \left(\frac{U_{rt}}{D_{rt}} - b_i \ln \frac{U_{rt}}{D_{rt}} \right) \qquad \text{(A.1)} \\ &+ (\gamma + \delta) b_i \ln \left(\frac{U_{rt}}{D_{rt}} \right) + \varepsilon_{irt} \end{aligned}$$

We estimate equation (A.1) and then test $\gamma + \delta = 0$. Notice that the estimation is replicated 100 times and the t-statistic reflects both the sampling variability and the uncertainty about the imputed values. For 2SLS model, the instruments are also rearranged accordingly so as to mirror the reparametrization of the endogenous variables. In particular, we instrument the term $(\ln \frac{U_{rt}}{D_{rt}} - b_i \ln \frac{U_{rt}}{D_{rt}})$ with $(Z_{r,t} - b_i Z_{r,t})$, while $b_i \ln \frac{U_{rt}}{D_{rt}}$ with $b_i Z_{r,t}$.

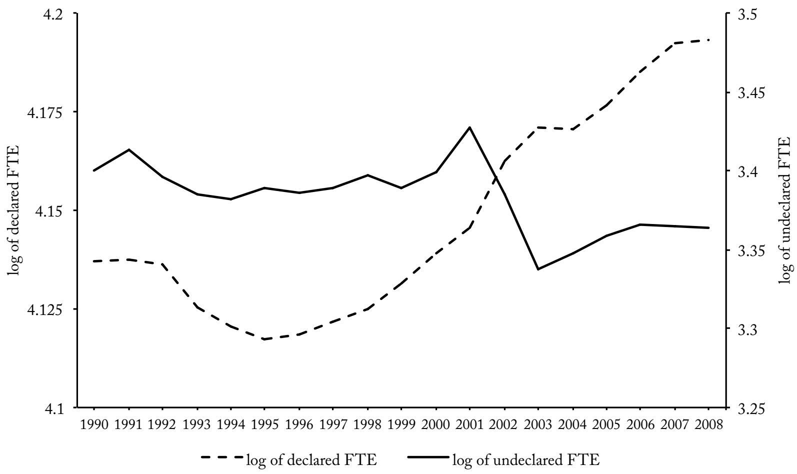

Figure 1: Evolution of declared and undeclared full time equivalent workers, 1990-2008.

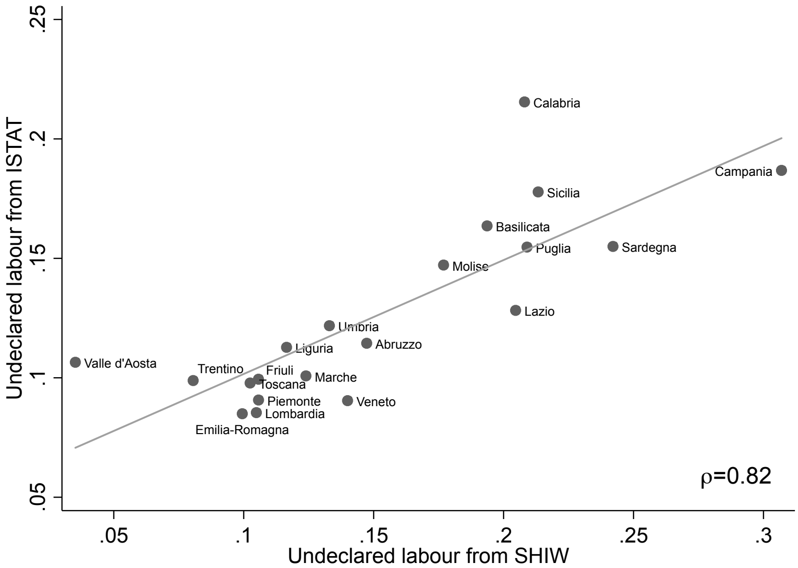

Figure 2: ISTAT's and SHIW's informality rate, 1995-2005.
Note: Data from ISTAT are for 1995, 1998, 2000, 2002, 2004 and 2005. Waves 1995, 1998, 2000, 2002, 2004 and 2006 for SHIW.

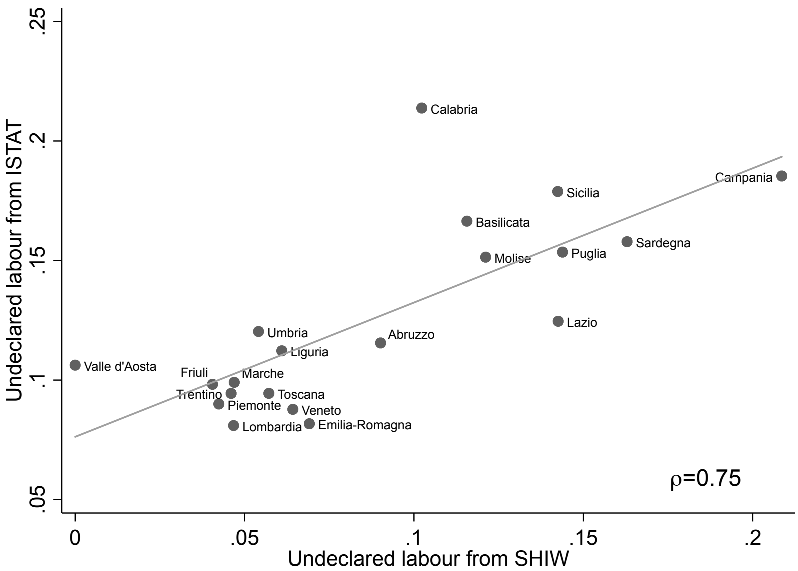

Figure 3: ISTAT informality rate and SHIW informality rate for dependent workers, 1995-2005.
Note: Data from ISTAT are for 1995, 1998, 2000, 2002, 2004 and 2005. Waves 1995, 1998, 2000, 2002, 2004 and 2006 for SHIW.

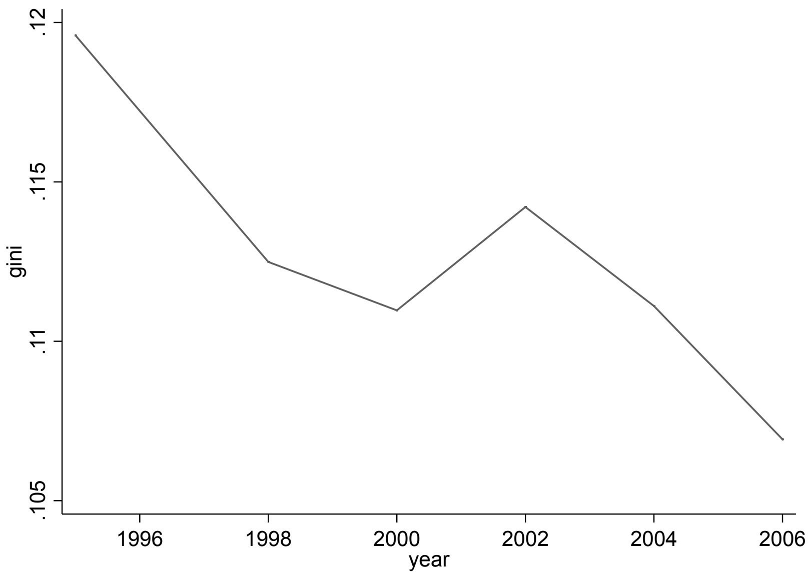

Figure 4: Gini index of hourly wage in SHIW, 1995-2006.

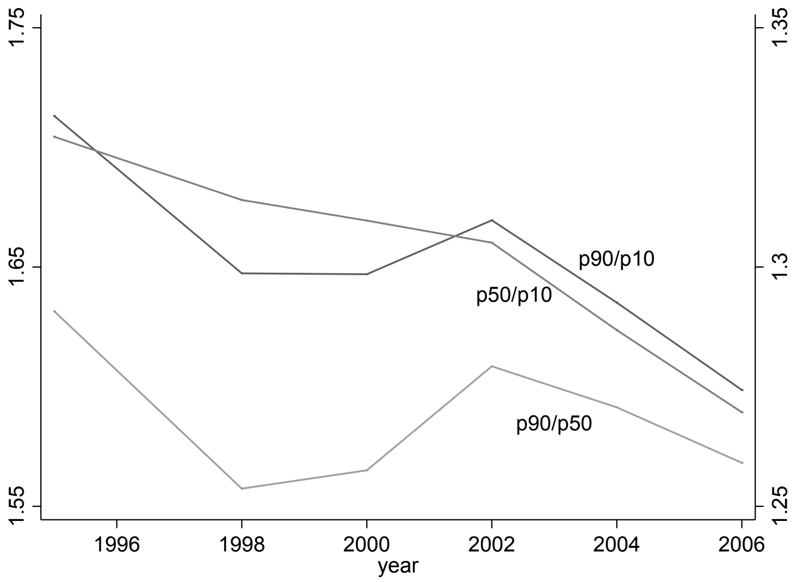

Figure 5: 90/10, 90/50 and 50/10 hourly wage inequality from SHIW, 1995-2006.
Note: 90/10 hourly wage inequality is measured along the left vertical axis; 90/50 and 50/10 indexes are represented along the right vertical axis.

Table 1: Description of the data.

	Undeclared status		
	Declared	**Undeclared**	**Total**
	%	%	%
Age			
15-24	6.8	17.8	7.8
25-39	42.0	50.6	42.7
40-59	49.4	30.6	47.7
60+	1.8	1.0	1.8
Sex			
Male	59.5	59.5	59.5
Female	40.5	40.5	40.5
Citizenship status			
Native	95.8	87.7	95.1
Immigrant	4.2	12.3	4.9
Marital status			
Married	64.3	41.9	62.3
Single	28.9	51.2	30.9
Separated	5.5	6.4	5.6
Widow(er)	1.3	0.5	1.2
School completed			
Less than high school	40.1	53.8	41.3
High school	47.2	34.8	46.1
College+	12.7	11.4	12.6
Potential work experience			
1-10	21.0	33.2	22.0
11-20	30.1	32.9	30.3
21-30	30.1	21.3	29.4
31+	18.8	12.6	18.3
Economic sector			
Agriculture	4.7	6.9	4.9
Manufacturing	31.1	20.9	30.2
Construction	5.8	12.5	6.4
Trade	11.6	18.5	12.2
Transport	3.9	4.1	3.9
Finance	4.0	1.9	3.8
Services	7.7	15.6	8.4
Public Administration, Defense, Education, Health care	31.3	19.5	30.3
Firm size			
1-19 employees	30.1	56.0	32.4
20-99 employees	24.4	20.6	24.0
100-499 employees	9.9	4.0	9.4
500+	10.5	3.7	9.9
Location			
North	54.0	33.0	52.2
Center	21.4	22.3	21.5
South	24.6	44.8	26.4
Hourly wage			
1st quartile	23.0	45.7	25.0
2nd quartile	25.5	22.7	25.2
3rd quartile	26.2	15.5	25.2
4th quartile	25.4	16.0	24.6
Mean of **hourly wage**	2.09	1.80	2.06
Stdev of **hourly wage**	(0.40)	(0.52)	(0.42)

The results are based on 11965 observations and weighted by sample weights.

Table 2: AR1 models for 20 NUTS2 regions.

Region	β	(s.e.)	R^2	RMSE
Piemonte	0.322	(0.459)	0.110	0.046
Valle d'Aosta	0.117	(1.885)	0.001	0.196
Lombardia	0.894	(0.155)	0.800	0.007
Trentino	0.198	(1.233)	0.006	0.193
Veneto	-0.445	(0.595)	0.123	0.048
Friuli	0.613	(0.400)	0.370	0.067
Liguria	0.246	(0.429)	0.076	0.032
Emilia Romagna	-0.255	(0.950)	0.018	0.038
Toscana	-0.318	(0.780)	0.040	0.084
Umbria	0.510	(0.432)	0.259	0.094
Marche	0.526	(0.376)	0.328	0.069
Lazio	-0.334	(0.694)	0.055	0.054
Abruzzo	0.683	(0.359)	0.475	0.041
Molise	0.821	(0.310)	0.636	0.063
Campania	0.474	(0.381)	0.279	0.045
Puglia	0.207	(0.485)	0.044	0.036
Basilicata	0.747	(0.274)	0.651	0.076
Calabria	-0.774	(0.404)	0.479	0.025
Sicilia	0.706	(0.152)	0.844	0.026
Sardegna	0.507	(0.228)	0.551	0.042

Table 3: Ordinary least squares estimates of own- and cross-wage labour demand elasticities.

	I	II	III	IV	V
$\log(\frac{U}{D})$	0.18***	0.19***	0.14**	0.31***	0.31***
	(0.06)	(0.06)	(0.06)	(0.08)	(0.09)
$\log(\frac{U}{D})$*Undeclared	-0.17**	-0.25**	-0.24**	-0.24**	-.24**
	(0.08)	(0.10)	(0.10)	(0.09)	(0.09)
Individual covariates	no	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes
Covariates at regional level	no	no	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	no	no	yes
No. of observations	5309		5289		

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. Model I controls for time effects and regional fixed effects. The models in columns II and III include a set of human capital and social characteristics, all interacted with an undeclared worker dummy. Models III, IV and V add interactions of the covariates with a year 2004 dummy. Model IV allows for the 2000 regional share of undeclared migrants, 2000 regional share of undeclared young workers, 2000 regional average education, 2000 regional share of undeclared workers in agriculture, construction and manufacturing, all interacted with a 2004 dummy. Model V also allows for lagged regional mean undeclared wages and the 2000 regional share of undeclared workers in the public sector. All models use sample weights. Sample: male workers employed in the private sector.

Table 4: Two stage least squares estimates of own- and cross-wage labour demand elasticities.

	I	II	III	IV	V
$\log(\frac{U}{D})$	0.108	0.064	0.020	0.356***	0.381***
	(0.151)	(0.140)	(0.141)	(0.112)	(0.104)
$\log(\frac{U}{D})$*Undeclared	-0.404**	-0.431**	-0.415**	-0.412**	-0.416**
	(0.190)	(0.180)	(0.167)	(0.168)	(0.169)
Individual covariates	no	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes
Covariates at regional level	no	no	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	no	no	yes
$\sigma_U = 1/\gamma + \delta$	-3.3	-2.7	-2.5	-18	-28
$\sigma_{DU} = 1/\delta$	-2.5	-2.3	-2.4	-2.4	-2.4
p-value H_0: $\gamma + \delta = 0$	0.205	0.100	0.070	0.760	0.840
			First-stage diagnostics		
Angrist and Pischke χ^2 (p-value)	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.000
Angrist and Pischke F-stat	46; 12	46; 17	50; 17	38; 17	47; 17
Stock and Yogo c.v. - 10% maximal IV size			16		
No of observations	5309		5289		

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. Relative supply of undeclared labour is instrumented by the difference between predicted and actual FTE undeclared workers multiplied by a 2004 dummy. Model I controls for time effects and regional fixed effects. Models in columns II and III include a set of human capital and social characteristics , all interacted with an undeclared worker dummy. Models III, IV and V add interactions of the covariates with a year 2004 dummy. Model IV allows for the 2000 regional share of undeclared migrants, 2000 regional share of undeclared young workers, 2000 regional average education, 2000 regional share of undeclared workers in agriculture, construction and manufacturing, all interacted with a 2004 dummy. Model V also allows for lagged regional mean undeclared wages and the 2000 regional share of undeclared workers in the public sector. All models use sample weights. Sample: male workers employed in the private sector.

Table 5: Two stage least squares estimates of the impact of undeclared work on declared wage inequality.

	I	II	III	IV	V	VI
			High school-Less than high school			
$\log(\frac{U}{D})$	0.384**	0.530**	0.737***	0.064	0.416**	0.460***
	(0.185)	(0.240)	(0.225)	(0.181)	(0.170)	(0.146)
$\log(\frac{U}{D})*$High school	0.367	0.382	0.400	0.263	0.265*	0.264*
	(0.240)	(0.244)	(0.245)	(0.151)	(0.152)	(0.152)
$\log(\frac{HS}{LHS})$	0.143**	0.170*	0.245**			
	(0.072)	(0.100)	(0.105)			
$\log(\frac{HS}{LHS})$*High school	0.197	0.207	0.210			
	(0.130)	(0.132)	(0.133)			
Individual covariates	yes	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-hs}/\sigma_{u-lhs}=\pi+\lambda/\pi$	2	1.7	1.5	4	1.6	1.55
p-value H_0: $\pi+\lambda=0$	0.000	0.000	0.000	0.098	0.000	0.000
			First-stage diagnostics			
Angrist and Pischke χ^2 (p-value)	0.000; 0.003	0.000; 0.003	0.000; 0.003	0.000; 0.000	0.000; 0.000	0.000; 0.000
Angrist and Pischke F-stat	24; 9	33; 8	34; 8	47; 14	37; 14	45; 14
Stock and Yogo c.v. - 10% maximal IV size				16		
No of observations				4425		

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. The relative supply of undeclared labor is instrumented by the difference between predicted and actual FTE undeclared workers multiplied by a 2004 dummy. All models include a set of human capital and social characteristics, all interacted with a high school dummy and a year 2004 dummy. All models use sample weights.

Table 6: Two stage least squares estimates of the impact of undeclared work on declared wage inequality.

	I	II	III	IV	V	VI
			College-High school			
$\log(\frac{U}{D})$	-0.263	0.348**	0.326**	-0.210	0.366**	0.320**
	(0.240)	(0.163)	(0.152)	(0.230)	(0.162)	(0.150)
$\log(\frac{U}{D})*$College	-0.251	-0.236	-0.233	-0.283	-0.270	-0.270
	(0.348)	(0.350)	(0.346)	(0.444)	(0.444)	(0.443)
$\log(\frac{CLG}{HS})$	-0.080*	-0.032	-0.050			
	(0.042)	(0.035)	(0.035)			
$\log(\frac{CLG}{HS})*$College	-0.055	- 0.050	-0.054			
	(0.170)	(0.170)	(0.174)			
Individual covariates	yes	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-clg}/\sigma_{u-hs}=\pi+\lambda/\pi$	1.9	0.3	0.3	2.34	0.26	0.15
p-value H_0: $\pi+\lambda=0$	0.183	0.782	0.825	0.276	0.847	0.916
			First-stage diagnostics			
Angrist and Pischke χ^2 (p-value)	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.005	0.000; 0.005	0.000; 0.000
Angrist and Pischke F-stat	30; 16	30; 16	63; 16	40; 8	35; 7	46; 7
Stock and Yogo c.v. - 10% maximal IV size				16		
No of observations				2377		

$***$ Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. The relative supply of undeclared labour is instrumented by the difference between predicted and actual FTE undeclared workers multiplied by a 2004 dummy. All models include a set of human capital and social characteristics, all interacted with a college dummy and a year 2004 dummy. All models use sample weights.

Table 7: Two stage least squares estimates of own- and cross-wage labour demand elasticities. Models with imputed gross wages.

	I	II	III	IV	V
$\ln(\frac{U}{D})$	0.084	0.041	0.018	0.393**	0.413***
	(0.180)	(0.162)	(0.160)	(0.134)	(0.125)
$\ln(\frac{U}{D})$*Undeclared	-0.324*	-0.351**	-0.350**	-0.342**	-0.350**
	(0.171)	(0.167)	(0.160)	(0.160)	(0.160)
Individual covariates	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes
Covariates at regional level	no	no	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	no	no	yes
$\sigma_U = 1/\gamma + \delta$	-4	-3.2	-3	19	16
$\sigma_{DU} = 1/\delta$	-3.1	-2.8	-2.8	-3	-2.8
p-value H_0: $\gamma + \delta = 0$	0.345	0.338	0.144	0.790	0.712
No of observations	5309			5289	

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Coefficients and standard errors are obtained by averaging over 100 separate models. Standard errors reflect sampling variability as well as uncertainty about the imputed values. All models use sample weights. For further information on the models' specifications, see note to Table 4.

Table 8: Two stage least squares estimates of the impact of undeclared work on declared wage inequality. Models with imputed gross wages.

	I	II	III	IV	V	VI
			High school-Less than high school			
$\ln(\frac{U}{D})$	0.440**	0.605**	0.800***	0.067	0.483***	0.520***
	(0.204)	(0.250)	(0.240)	(0.203)	(0.190)	(0.165)
$\ln(\frac{U}{D})$*High school	0.301	0.320	0.322	0.224	0.230	0.226
	(0.228)	(0.232)	(0.233)	(0.150)	(0.150)	(0.150)
Controlling for HS/LHS supply	yes	yes	yes	no	no	no
Individual covariates	no	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-hs}/\sigma_{u-lhs}=\pi+\lambda/\pi$	1.7	1.5	1.4	4.3	1.4	1.4
p-value H_0: $\pi+\lambda=0$	0.000	0.000	0.000	0.194	0.000	0.000
No of observations			4425			
			College-High school			
$\ln(\frac{U}{D})$	-0.255	0.370*	-0.332*	-0.200	0.385*	0.330*
	(0.262)	(0.22)	(0.195)	(0.25)	(0.213)	(0.195)
$\ln(\frac{U}{D})$*College	-0.305	-0.290	-0.286	-0.350	-0.335	-0.335
	(0.370)	(0.371)	(0.370)	(0.473)	(0.475)	(0.474)
Controlling for CLG/HS supply	yes	yes	yes	no	no	no
Individual covariates	yes	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-clg}/\sigma_{u-hs}=\pi+\lambda/\pi$	2.4	0.21	1.9	2.7	0.13	-0.01
p-value H_0: $\pi+\lambda=0$	0.197	0.862	0.920	0.277	0.363	0.997
No of observations			2377			

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Coefficients and standard errors are obtained by averaging over 100 separate models. Standard errors reflect sampling variability as well as uncertainty about the imputed values. All models use sample weights. For further information on the models' specifications, see note to Tables 5 and 6.

Table 9: Robustness 1: two stage least squares estimates of own- and cross-wage labour demand elasticities. Male and female workers in the public and the private sectors.

	Males and Females		Private and public sector	
	I	II	III	IV
$\log(\frac{U}{D})$	0.301***	0.250**	0.160**	0.104**
	(0.101)	(0.100)	(0.065)	(0.054)
$\log(\frac{U}{D})$*Undeclared	-0.352**	-0.352**	-0.363***	-0.364***
	(0.145)	(0.145)	(0.123)	(0.124)
Individual covariates	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes
Covariates at regional level	yes	yes	yes	yes
Lagged regional undeclared wages, share in public sector	no	yes	no	yes
$\sigma_U = 1/\gamma + \delta$	-24	-11	-5	-3.8
$\sigma_{DU} = 1/\delta$	-2.8	-2.8	-2.8	-2.8
p-value H_0: $\gamma + \delta = 0$	0.736	0.503	0.064	0.021
First-stage diagnostics				
Angrist and Pischke χ^2 (p-value)	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.000
Angrist and Pischke F-stat	36; 16	45; 16	37; 17	46; 18
Stock and Yogo c.v. - 10% maximal IV size	16			
No of observations	8222		11965	

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. All models use sample weights. For further information on the models' specifications, see note to Table 4.

Table 10: Robustness 2: two stage least squares estimates of the impact of undeclared work on declared wage inequality. Male and female workers in the public and the private sector.

	I	II	III	IV
HIGH SCHOOL-LESS THAN HIGH SCHOOL				
$\log(\frac{U}{D})$	0.440**	0.192*	0.290**	0.096
	(0.215)	(0.117)	(0.132)	(0.078)
$\log(\frac{U}{D})*$High school	0.366	0.308	0.244*	0.237*
	(0.230)	(0.196)	(0.142)	(0.127)
Controlling for HS/LHS supply	yes	yes	no	no
Individual covariates	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes
Covariates at regional level	yes	yes	yes	yes
Lagged regional undeclared wages, share in public sector	yes	yes	yes	yes
$\sigma_{u-hs}/\sigma_{u-lhs}=\pi+\lambda/\pi$	1.8	2.6	1.8	3.5
p-value H_0: $\pi+\lambda=0$	0.000	0.000	0.001	0.002
FIRST-STAGE DIAGNOSTICS				
Angrist and Pischke χ^2 (p-value)	0.000; 0.004	0.000; 0.003	0.000; 0.000	0.000; 0.000
Angrist and Pischke F-stat	33; 8	35; 8	37; 17	45; 15
Stock and Yogo c.v. - 10% maximal IV size	16			
No of observations	6819	9398	6819	9398
COLLEGE-HIGH SCHOOL				
$\log(\frac{U}{D})$	0.289***	0.250***	0.310***	0.250***
	(0.110)	(0.091)	(0.120)	(0.094)
$\log(\frac{U}{D})*$College	-0.093	-0.173	-0.078	-0.170
	(0.235)	(0.172)	(0.275)	(0.186)
Controlling for CLG/HS supply	yes	yes	no	no
Individual covariates	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes
Covariates at regional level	yes	yes	yes	yes
Lagged regional undeclared wages, share in public sector	yes	yes	yes	yes
$\sigma_{u-clg}/\sigma_{u-hs}=\pi+\lambda/\pi$	0.6	0.3	0.7	0.3
p-value H_0: $\pi+\lambda=0$	0.500	0.700	0.476	0.720
FIRST-STAGE DIAGNOSTICS				
Angrist and Pischke χ^2 (p-value)	0.000; 0.004	0.000; 0.000	0.000; 0.001	0.000; 0.000
Angrist and Pischke F-stat	65; 17	53; 17	45; 10	45; 13
Stock and Yogo c.v. - 10% maximal IV size	16			
No of observations	3916	6686	3916	6686

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. Models I and III include female workers; models II and IV bring also in public sector employees. All models use sample weights. For further information on the models' specifications, see note to Tables 5 and 6.

Table 11: Robustness 3: two stage least squares estimates of own- and cross-wage labour demand elasticities. Conservative definition of undeclared worker.

	I	II	III	IV	V
$\log(\frac{U}{D})$	0.171	0.120	0.067	0.230**	0.100
	(0.155)	(0.145)	(0.146)	(0.115)	(0.135)
$\log(\frac{U}{D})$*Undeclared	-0.671*	-0.571**	-0.554**	-0.542**	-0.540**
	(0.377)	(0.230)	(0.221)	(0.223)	(0.224)
Individual covariates	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes
Covariates at regional level	no	no	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	no	no	yes
$\sigma_U = 1/\gamma + \delta$	-2	-2.2	-2	-3.2	-2.2
$\sigma_{DU} = 1/\delta$	-1.5	-1.7	-1.8	-1.8	-1.8
p-value H_0: $\gamma + \delta = 0$	0.180	0.066	0.050	0.135	0.033
First-stage diagnostics					
Angrist and Pischke χ^2 (p-value)	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.000	0.000; 0.000
Angrist and Pischke F-stat	46; 15	46; 23	51; 24	79; 23	59; 23
Stock and Yogo c.v. - 10% maximal IV size			16		
No of observations	5309	5289			

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. All models use sample weights. Sample: male workers employed in the private sector. For further information on the models' specifications, see note to Table 4.

Table 12: Robustness 4: two stage least squares estimates of the impact of undeclared work on declared wage inequality. Conservative definition of undeclared worker.

	I	II	III	IV	V	VI
HIGH SCHOOL-LESS THAN HIGH SCHOOL						
$\log(\frac{U}{D})$	0.422**	0.466**	0.510**	0.122	0.164	0.093
	(0.180)	(0.204)	(0.182)	(0.162)	(0.151)	(0.157)
$\log(\frac{U}{D})$*High school	0.390	0.400	0.400	0.285	0.290*	0.300*
	(0.270)	(0.271)	(0.271)	(0.170)	(0.171)	(0.172)
Controlling for CLG/HS supply	yes	yes	yes	no	no	no
Individual covariates	yes	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-hs}/\sigma_{u-lhs}$=$\pi+\lambda/\pi$	1.9	1.8	1.7	3.3	2.7	4.2
p-value H_0: $\pi+\lambda=0$	0.000	0.000	0.000	0.032	0.005	0.036
FIRST-STAGE DIAGNOSTICS						
Angrist and Pischke χ^2 (p-value)	0.000; 0.003	0.000; 0.003	0.000; 0.003	0.000; 0.003	0.000; 0.003	0.000; 0.003
Angrist and Pischke F-stat	24; 8	102; 8	89; 8	48; 14	77; 14	59; 14
Stock and Yogo c.v. - 10% maximal IV size	16					
No of observations	4665					
COLLEGE-HIGH SCHOOL						
$\log(\frac{U}{D})$	-0.105	0.350**	0.030	-0.045	0.420***	0.260*
	(0.204)	(0.166)	(0.334)	(0.215)	(0.142)	(0.141)
$\log(\frac{U}{D})$*College	-0.300	-0.270	-0.273	-0.331	-0.312	-0.313
	(0.371)	(0.371)	(0.370)	(0.474)	(0.472)	(0.470)
Controlling for CLG/HS supply	yes	yes	yes	no	no	no
Individual covariates	yes	yes	yes	yes	yes	yes
Region fixed effects	yes	yes	yes	yes	yes	yes
Covariates at regional level	no	yes	yes	no	yes	yes
Lagged regional undeclared wages, share in public sector	no	no	yes	no	no	yes
$\sigma_{u-clg}/\sigma_{u-hs}$=$\pi+\lambda/\pi$	3.8	0.2	-8	8.3	-0.25	-0.2
p-value H_0: $\pi+\lambda=0$	0.256	0.816	0.454	0.380	0.814	0.900
FIRST-STAGE DIAGNOSTICS						
Angrist and Pischke χ^2 (p-value)	0.000; 0.000	0.000; 0.000	0.000; 0.001	0.000; 0.005	0.000; 0.005	0.000; 0.005
Angrist and Pischke F-stat	29; 16	90; 16	36; 16	40; 8	83; 7	63; 8
Stock and Yogo c.v. - 10% maximal IV size	16					
No of observations	2451					

*** Significant at 1%, ** Significant at 5%, * Significant at 10%. Standard errors (in parentheses) account for clustering on region and year of observation. All models use sample weights. For further information on models' specifications, see note to Tables 5 and 6.