CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 786

The Perverse Effect of Flexible Work Arrangements on Informality

Edoardo Di Porto, Pietro Garibaldi, Giovanni Mastrobuoni, and Paolo Naticchioni

July 2026

University of Naples Federico II

University of Salerno

Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS AND STATISTICS – UNIVERSITY OF NAPLES FEDERICO II
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unina.it
ISSN: 2240-9696

WORKING PAPER NO. 786

The Perverse Effect of Flexible Work Arrangements on Informality

Edoardo Di Porto*, Pietro Garibaldi†, Giovanni Mastrobuoni‡, and Paolo Naticchioni§

Abstract
Work arrangements with no guaranteed working hours, which include casual work, are often promoted as a means to regularise informal labour. Utilising unique Italian administrative data that links employer-employee records, daily voucher usage by firms, and randomly timed labour inspections (2014-2017), we demonstrate that this type of Flexible Work Arrangements (FWA) can also hinder enforcement and increase undeclared work. We document that, upon inspection, some firms validate undeclared work with FWAs on the spot, raising the probability of FWA usage by 0.88 percentage points (18%) on average, with the largest increases occurring on the day of and the day after the inspection. A simple partial-equilibrium labour-demand model with heterogeneous tax morale rationalises these "on-the-spot" validations as an enforcement-avoidance margin. The post-inspection increase vanishes when firms are required to pre-notify the tax authority of their use of FWAs. Moreover, when FWAs are completely abolished, presumptive misusers substitute FWAs with temporary contracts.

JEL Classification: J23, H26.

Keywords: informality, labour vouchers, flexible work arrangements, occasional work, zero-hour contracts.

Acknowledgments: We thank Bruno Anastasia, Saverio Bombelli, Lorenzo Cappellari, Tullio Jappelli, Marteen Lindeboom, Roberto Nisticò, Lorenzo Pandolfi, Matteo Paradisi, Annalisa Scognamiglio, and participants at conferences and seminars at AIEL, Bank of Italy, Inps, CSEF, Mannheim, Georgian Economic Association, University of Venice, IPEA-Brasilia, University of Buenos Aires. We would also like to thank the Direzione Centrale Studi e Ricerche as well as the Direzione Centrale Entrate of the Italian Social Security Administration (INPS), for their invaluable support with the data and the institutional setting, as well as the sponsors of the VisitINPS program for their financial support. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS. All errors are our own.

* Sapienza University of Rome, CSEF, University of Naples "Federico II", UCFS, Uppsala University, and CESifo. E-mail: edoardo.diporto@uniroma1.it

† Collegio Carlo Alberto, University of Torino, CEPR, and IZA. E-mail: pietro.garibaldi@unito.it

‡ Collegio Carlo Alberto, University of Torino, CEPR, IZA, and RFBerlin. (*Corresponding author*) E-mail: giovanni.mastrobuoni@carloalberto.org

§ Roma Tre University and IZA. E-mail: paolo.naticchioni@uniroma3.it

1 Introduction

In most advanced economies, the underground economy is estimated to account for 10-20% of GDP (Schneider and Enste, 2002).[1] Furthermore, undeclared work is notably persistent and, if anything, its share has been estimated to have increased over time (Ulyssea, 2018).

Partly in response to shadow work, many governments have introduced more flexible labour contracts, known as flexible work arrangements (FWAs) (ILO, 2013).[2] Most European countries have arrangements with no guaranteed working hours, allowing firms to swiftly adjust labour demand while offering workers more adaptable schedules.[3] Boeri et al. (2020) provide a recent survey of the small but expanding economic literature on flexible arrangements.[4]

Although there is little causal evidence regarding the relationship between FWAs and undeclared work, many experts believe that this relationship should be negative.[5]

Among others, the *European Union Agency for the Improvement of Living and Working Conditions* mentions that these new forms of employment have been developed to help formalise undeclared work practices.[6] The *European Platform Tackling Undeclared Work* has stated that FWAs, in the form of labour vouchers (government-backed payment instruments that households or firms can use to pay for casual yet legal work), should target areas where undeclared work is prevalent (see Williams, 2018).

This study challenges the notion that FWAs inherently reduce undeclared work. While they

[1]In developing countries the share is closer to 50% (Schneider and Enste, 2000).

[2]For example, according to the ILO, Belgium, France, Finland, Denmark, and Switzerland, introduced FWAs to discourage undeclared work.

[3]Chen et al. (2019) estimate that Uber drivers, possibly due to selection, benefit enormously from real-time flexibility, and Chan (2018) shows that emergency department physician work schedules can distort effort allocation and patient care.

[4]Katz and Krueger (2019) cover FWAs for the US and Adams and Prassl (2018) and Datta et al. (2019) for Europe.

[5]There is also evidence that rigid employment protection legislation reduces job flows and forces firms to hire workers with more temporary contracts or off the books (for an early and a more recent survey, see Schneider and Enste, 2000, Ulyssea, 2020). Theoretical and empirical contributions that show that labour market rigidities increase informality include Blanchard and Portugal (2001), Fugazza and Jacques (2004), Albrecht et al. (2009b), Maloney (2004), Johnson et al. (1998), and (DiPorto et al., 2017).

[6]See https://www.eurofound.europa.eu/topic/undeclared-work.

offer flexibility, we demonstrate that such arrangements can hinder labour inspectors' ability to identify undeclared work. When misused, FWAs, such as the Italian labour vouchers examined here, can enable firms to legitimise the physical presence of workers engaged in undeclared employment.

An additional contribution of this paper is to document under-declared work as a form of partial informality: workers may hold a formal contract while part of their true labour input (and remuneration) remains off the books, consistent with a handful of studies on wage under-reporting and envelope payments in the economics literature (Biro et al., 2022, Gavoille and Zasova, 2023, Tonin, 2011).

Similar work arrangements exist in several countries, for example, UK's "Zero Contract Hours," where workers may work more than the officially declared number of hours, or Italy's labour vouchers, where for a day of work a worker may receive a single voucher for his/her work, so as to justify his/her physical presence in the workplace, and be paid the rest under the table. The Italian Ministry of Labour in 2016 realised that some firms may strategically comply with voucher requirements only when faced with an imminent inspection, similar to "a passenger validating a ticket only when the inspector arrives."[7] The Ministry report also provides evidence that about 10% of FWA workers were previously employed by the same firm but with regular contracts.

Consequently, we develop a model of firms' labour demand across contract types, allowing for "spot validation" upon inspection. Firms differ in tax morale. In the absence of on-the-spot regularisation, more flexible legal contracts reduce the incentive to engage in undeclared work.[8] With spot validation, however, low-tax-morale firms can use vouchers as an inspection-contingent regularisation device, while tax-abiding firms' behaviour is largely unchanged.

[7] See Ministero del Lavoro (2016).

[8] This is in line with, among others, Albrecht et al. (2009a), Bosch and Esteban-Pretel (2012), and Ulyssea (2018).

A central implication is that enforcement and institutional design interact. Stronger enforcement (a higher inspection probability) reduces shadow activity, but this deterrent effect is weaker when on-the-spot regularisation is easier. Intuitively, if firms can quickly relabel undeclared jobs as vouchers at inspection, the expected punishment from inspection falls, and firms increase their use of shadow employment.

We bring these predictions to the data in Italy, a country with a sufficiently large underground economy and legislation that has first liberalised and later abolished FWAs. Within the European Union, only some Eastern European countries, as well as Spain and Greece, exhibit a higher prevalence of undeclared work[9] and, beginning about 15 years ago, employers in Italy could purchase 10-euro vouchers to pay for work without needing a standard labour contract. The worker would later exchange vouchers for money. This system was intended to discourage undeclared work by eliminating bureaucracy and reducing hiring and firing costs.[10]

Labour force surveys often lack detail to identify FWA arrangements (Katz and Krueger, 2019), and when paid sums do not contribute to future social security benefits, even administrative data can be uninformative. This is not the case in Italy, where government-backed vouchers are fully recorded.[11]

We use a unique data set drawn from three separate Italian administrative records: i) employer-employee social security records that cover the period 2014-2017; ii) daily firm-level usage of vouchers between 2014 and 2017; and finally, iii) data on all labour inspections between 2014 and 2017. Our data indicate that almost one in four firms in 2016 used vouchers.

Exploiting documented unpredictability of the timing of labour inspections, we test the first

[9] According to Williams et al. (2017), 17.2% of Italian work in undeclared (the EU average is 16.4%).

[10] In Europe voucher-based work is also available in Austria, Belgium, Ireland, Finland, France, Hungary, Lithuania, and Slovenia (Mandl, 2020).

[11] Additionally, as shown by Mas and Pallais (2017), workers often select into FWA contracts, complicating the estimation of counterfactual scenarios. We discuss the identification in Section 5.1.

prediction of the theoretical model. We find clear evidence that–as soon as an inspection starts–some firms tend to immediately increase their use of vouchers. The probability of voucher use increases by 0.88 percentage points immediately after the onset of inspection, which corresponds to a relative increase of approximately 18%. The greatest changes occur on the day of inspection and the day after, respectively, 1.5 (30%) and 1.4 percentage points (29%). We also show, consistently with the second prediction of the model, that this post-inspection effects basically disappear when firms are required, in October 2016, to pre-notify, with at least one hour notice, the tax authority of their use of FWAs, making it harder to use vouchers for on-the-spot validation of shadow jobs.

Next, we test whether FWAs displace regular work. We refer to misusing firms as firms that exhibit an inspection-induced discrete increase in voucher use (consistent with inspection-contingent regularisation), and to compliant firms as firms that do not. We analyse what these two sets of firms did in March 2017 when vouchers were finally abolished.

This measure of misusing firms is subject to misclassification error, as changes in individual-level use of FWAs around labour inspections may reflect unobservable demand shocks. However, we show that our main estimates can be interpreted as lower bounds of the true effects. In addition, we argue that varying the definition of misuse, and thus the size of the effects, can be informative about the fraction of firms hiding undeclared work. We show that approximately 18% of the inspected firms that use vouchers misuse FWAs.

In line with the third prediction of the model, "misusing" firms are shown, when vouchers are abolished, to revert to the next most flexible work contracts, hiring approximately two additional fixed-term workers, representing a 50% increase with respect to the pre-abolition average of misusing firms. Due to these substitution effects, the total declared wage bill–including vouchers–remains unchanged. Finally, there is suggestive evidence that after March 2017, when firms can no longer hide undeclared work using vouchers, labour inspectors detect more evasion.

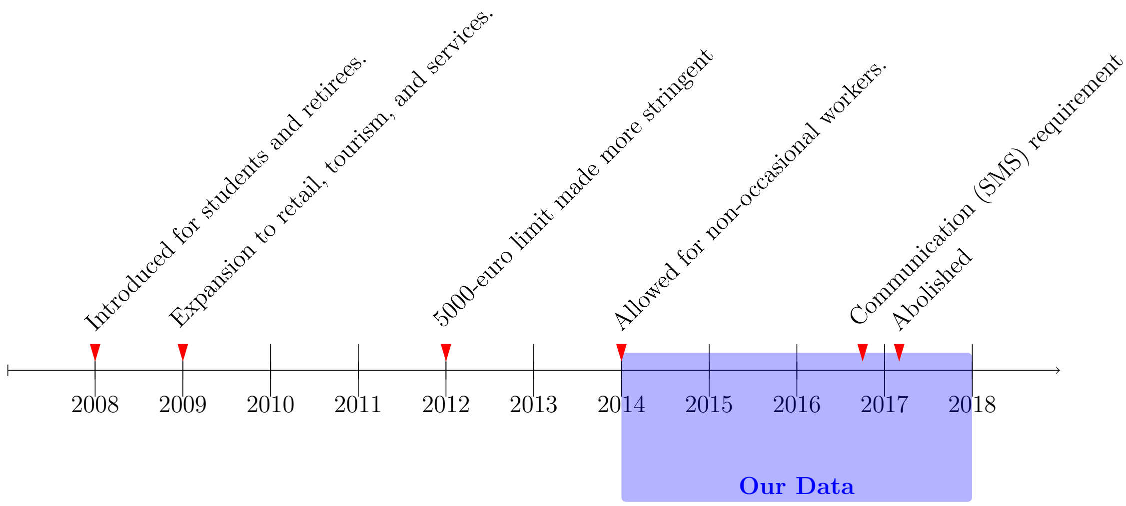
The study proceeds as follows. Section 2 describes the institutional setting. Section 3 presents and solves a simple labour demand model of labour vouchers and derives the optimal choice of contracts, with and without the option to go shadow. The section highlights the main empirical predictions. Section 4 describes the data set used in the paper. Sections 5 and 6 present the empirical models and the evidence, while Section 7 concludes.

2 Institutional framework

In this section, we describe the regulatory framework governing the two main institutions analysed in the paper: Italian labour vouchers and labour inspections.

Vouchers: In 2008, the Italian legislator introduced FWAs in the extreme form of labour vouchers. Figure 1 outlines the timeline from the first time FWAs were introduced to the time they were abolished. Employers could buy vouchers from the Social Security Administration (INPS), or in tobacco stores, banks, and post offices. The vouchers looked like cheques (see Appendix Figure A1), and employers who wanted to pay someone for completing a very short temporary job would fill them with the worker's social security number ("*Codice Fiscale*") and the date of the job/task. Once the worker was paid with the voucher, he/she could redeem it for cash. For every 10 euros paid by the employer, the worker received 7.50 euros, 1.30 euros covered the social security contributions, which is lower than the contribution rate under standard employment contracts (approximately one-third of gross pay), 70 cents for health insurance and 50 cents for the commission fee paid to the social security administration.

Figure 1: Timeline of Voucher Legislation



Initially, vouchers had considerable restrictions: private citizens could only spend a total of €5000 vouchers for each worker who helped them in day-to-day casual work (house cleaning, gardening, etc.), and employers faced an even lower threshold (€2500). Moreover, only students and retirees were allowed to receive vouchers, and only in the agricultural sector.

Several small changes in these restrictions led to a steep increase in the use of FWAs. Initially, the centre-right government extended the use of vouchers to all workers in the agricultural sector, not just students and retirees. More limitations were lifted in the following years and, as shown in Figure 2, this led to rapid growth in the monthly number of 10-euro vouchers sold: from a few thousand in 2008 to a peak of almost 20 million in 2016.

In 2009, vouchers became available in the retail, tourism, and service, and for domestic workers. One year later, they were fully liberalised, extended to all sectors and all categories of workers. In 2012, the worker's €5000 limit was made more stringent, as it applied to the sum across all the employers and not to each employer separately, while a new limit was introduced: each firm could at most spend €2000 in vouchers on any single worker. In 2015, the labour reform (called the "Jobs Act") allowed vouchers not to be related to occasional

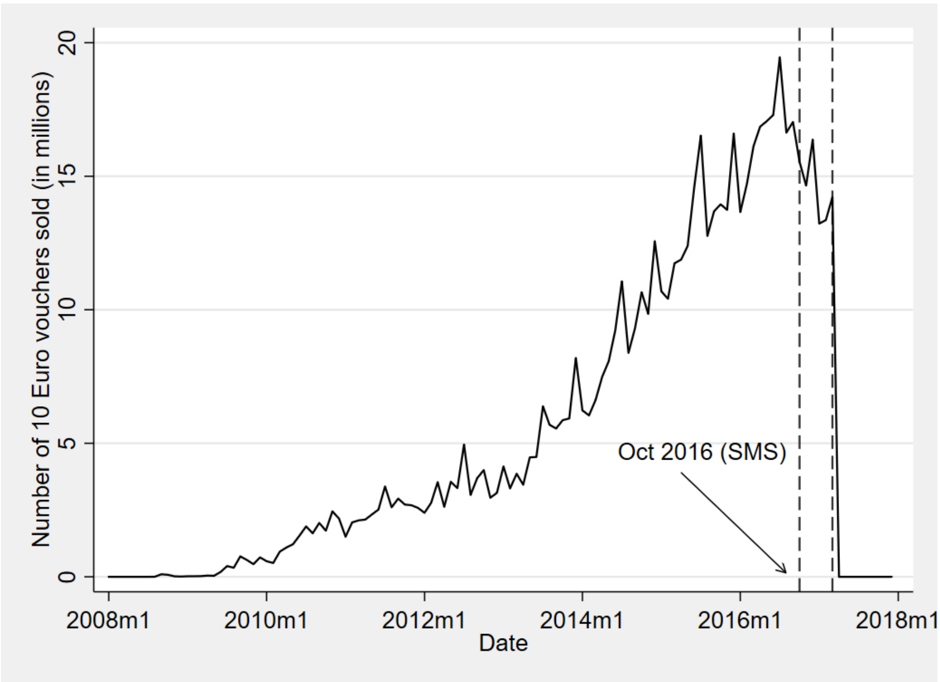


Figure 2: Vouchers Sold

Notes: The figure plots the monthly total number of 10-euro vouchers sold.

work, and their annual limit increased to €7000 (leaving the amount firms could use on a single worker unchanged).

The use of vouchers reached a peak in 2016, when the pressure from labour unions to reform their use or completely abolish them intensified. In October 2016, a first reform was passed, and employers had to inform the Social Security Administration by text message at least 60 minutes before using a voucher. Several months later, as pressure increased and a new government was in power, the use of vouchers was outlawed.[12]

Labour Inspections: Inspections are conducted without prior notice to maximise the chance of detecting any undeclared workers or any hazardous working conditions.[13] They are divided into three phases. First, each inspection is secretly planned to ensure that the second phase, the subsequent access to the firm's premises, is carried out quickly by a sufficient number of inspectors, so as to prevent workers from leaving the premises. The second phase can last up to several days and ends with a preliminary report. Inspectors collect accounting books and any other useful information and interview employers and employees. During the last phase, which can last at most 90 days from the notification of the preliminary report (Art. 14 of Law No. 689/1981), visits to the firm are rarer, and the inspectors prepare the final report.

From the firm's standpoint, inspections occur at random and, therefore, unpredictable times, and their duration ranges from a few days to several weeks. In order to deter any information leakage, corrupt inspectors who inform firms about an imminent inspection in exchange for money face up to 10 years of imprisonment (Article 319 of the Criminal Code). Depending on the outcome of the inspection, firms may have to pay a fine and/or rectify any irregularities before resuming production.

Before exploiting these institutional changes in the empirical analysis, we develop a labour

[12] A much more limited version of vouchers was reintroduced at the end of 2017.

[13] The rules can be found in the "Circolare INPS No. 76/2016."

demand model that generates potential links between labour demand, including labour vouchers, labour inspections, and the use of undeclared work.

3 A Labour Demand Model of Jobs, Temporary Jobs and FWAs

3.1 The Environment and the Institutions

Following the tradition of the search and matching literature and the Diamond-Mortensen-Pissarides model, we develop a stylised static labour demand model with heterogeneous firms characterised by a single job. The formal modelling of labour contracts is in line with Cahuc et al. (2016). There is a measure 1 of homogeneous workers and an endogenous number of heterogeneous firms (jobs). Firm heterogeneity is defined by its technological probability of becoming unproductive. Each job (or firm) faces a probability λ of becoming unproductive, where $\lambda \in [0, 1]$ is drawn from a cumulative probability distribution $G(\lambda)$ when the worker and the firm meet, and $G(1) = 1$.[14] A job produces output y for the fraction of time $1 - \lambda$, and produces 0 for the rest of the time (normalised to be one). After observing λ, the firm decides which type of contract to offer to a single worker. Contracts differ by their termination costs.

The wage ω paid to each worker is taken as given by the firm.[15] The labour market is characterised by a payroll tax τ, regardless of the type of contract. The tax is paid on a flow basis by the firm and, at first, we assume that the tax cannot be evaded.

[14]Cahuc et al. (2016) provide a matching model with heterogeneous jobs that features job heterogeneity and optimal contract selection. See Boeri and Garibaldi (2024) for a recent survey on the vast literature on fixed term contracts.

[15]To focus on how λ influences labour demand, we assume a fixed wage ω. A simple extension would be to assume that more flexible labour contracts come with lower wages. The model would be solved with an endogenous wage and rent sharing.

In Section 3.2 we introduce the possibility of tax evasion, allowing contracts to be shadow. When labour is shadow (or black), taxes are avoided, but firms risk being fined by the labour inspectors. In reality most firms avoid the use of shadow employment, therefore we add a further dimension of heterogeneity, a firm-specific "tax-morale" or "tax-evasion attitude" $\theta^j > 0$ associated with the status of being a tax dodger. For simplicity, we assume that θ^j takes only two values, $\theta^j \in \{\theta^l, \theta^h\}$ with $\theta^l < \theta^h$, and that a proportion p of firms have $\theta = \theta^h$. Conditional on the firm type θ, jobs draw λ from a type-specific distribution $G(\lambda \mid \theta)$. The unconditional distribution of λ in the economy is therefore $G(\lambda) = p\, G(\lambda \mid \theta^h) + (1-p)\, G(\lambda \mid \theta^l)$.

Without tax evasion, the main choice of the firm is about which contract to offer depending on λ. Broadly speaking, the Italian legislation allows three different types of labour contracts: open-ended, fixed-term jobs, and FWAs/vouchers (we use the words FWAs or vouchers interchangeably). Different contracts have different termination costs. When faced with an open-ended contract that is unproductive, the firm is better off paying a firing tax equal to $-F$. In line with Italian legislation, we assume that the tax is proportional to the wage rate: $F = f\omega$.[16] $J^i(\lambda)$ with $\{i = oe, ft, v\}$ represent the value of each job contract, where oe refers to open-ended, ft to fixed term, and v to vouchers.

Fixed-term contracts are active for a fraction $1 - \rho^{ft}$ of the time. When a firm opens a fixed-term contract, the worker produces and is going to be paid for an expected duration equal to $1 - \rho^{ft}$, regardless of the specific value of the job λ. The advantage of a fixed-term contract is that the firm does not pay any firing costs when the productivity drop strikes (which happens with probability λ^{ft}).[17] However, the downside is that the firm is obliged to pay the worker until the contract expires (which happens with probability ρ^{ft}) even if

[16] The model abstracts from the slight differences in social security contribution rates across contracts, but this should be second-order for our empirical analysis, as those differences do not vary over time with the arrival of labour inspectors.

[17] The value of the firing tax $F = f\omega$ has the restriction that $f < 1 + \frac{\tau}{\omega}$, so that the firing tax for open-ended contracts must be smaller than the firing tax paid under a fixed-term contract.

λ strikes and productivity drops to 0.[18] When the fixed term contract expires, the job ends, even if it is still productive. Finally, the firm can open FWAs. FWAs do not have any dismissal cost, but are characterised by an expected duration $1-\rho^v$, where ρ^v is considerably larger than ρ^{ft}.[19] In practice, it is as if FWAs can be terminated at any time at no cost.

There is a fixed cost of opening a job, any job, where legal or not, equal to K, which means that a firm will open a λ job as long as $J^i(\lambda) \geq K$ for at least one type of contract i. Conditional on the realised technological parameter λ, firms choose the type of contract to offer such that $J^*(\lambda) = \max_{\{i=[oe,ft,v]\}} \{J^i(\lambda)\}$, where $J^*(\lambda)$ is the optimal labour demand. Furthermore, jobs are created as long as $J^*(\lambda) \geq K$.

To solve this maximisation problem, the value functions of the three jobs J^{oe}, J^{ft}, J^v are:

$$J^{oe}(\lambda) = (1-\lambda)(y-\tau-\omega) - \lambda F, \tag{1}$$

$$J^{ft}(\lambda) = (1-\rho^{ft})\left[(1-\lambda)y - \omega - \tau\right], \tag{2}$$

$$J^{v}(\lambda) = (1-\rho^{v})\left[(1-\lambda)(y - \omega - \tau)\right]. \tag{3}$$

Since all job values are monotonically decreasing in λ, the maximisation problem satisfies the reservation property. In other words, maximisation is solved by selecting particular values of λ (say λ^{ft} and λ^v) such that each contract is optimal in any particular subset of the domain of G in $[0,1]$. Furthermore, one can easily show that $J^{oe}(0) > J^{ft}(0) > J^v(0)$ and that $J^v(1) = 0 > J^{ft}(1) = -(\omega+\tau) > J^{oe}(1) = -F$.

Proposition 1. *The value of the firms' optimal labour contract $J^*(\lambda)$ depends on two productivity thresholds λ^{ft}, λ^v such that $\forall \lambda < \lambda^{ft}$ firms choose an open-ended contract, $\forall \lambda \in [\lambda^{ft} - \lambda^v]$ firms choose a fixed-term contract, and $\forall \lambda > \lambda^v$ firms choose a flexible*

[18] The data confirm that only a handful of fixed-term contracts end with a dismissal (0.5 per cent in 2016), consistently with our assumption that no firing costs are paid to fixed-term contracts.

[19] Note that this is in line with the relatively low caps on the amounts firms are allowed to pay each worker in labour vouchers (see Section 2).

working arrangement. Moreover, firms do not open jobs $\forall \lambda \geq \lambda^e$ *where* $J(\lambda^e) = K$.

The maximisation is thus an envelope of three downward-sloping lines. The reservation probabilities λ^{ft} and λ^v make the firm indifferent between an open ended and a fixed term job and between a fixed term and a voucher: $J^{oe}(\lambda^{ft}) = J^{ft}(\lambda^{ft})$ and $J^{ft}(\lambda^v) = J^v(\lambda^v)$. The values of λ^{ft} and λ^v are specified in Appendix A. The intuition behind this result is as follows. For a given net flow productivity $\tilde{y} = y - \omega - \tau$, a firm of type λ has a strong ordering about which contract to offer. On one extreme, open-ended contracts are ideal for jobs with a very long expected duration, while, on the other extreme, FWAs are ideal for jobs that are expected to last very little. Thus, introducing FWAs creates opportunities for labour demand that would otherwise be absent. If the probability of meeting a worker is ξ, total employment is simply $n = \xi G(\lambda^e)$, while $(1 - \xi G(\lambda^e))$ is non-employment. Without loss of generality, we set $\xi = 1$.

3.2 Black and Grey Shadow Employment

We now introduce the possibility of evading taxes by under-reporting jobs completely ("black" shadow work), as opposed to hide such jobs behind the veil of vouchers ("grey" shadow work).

A black shadow job is an undeclared employment relationship that does not comply with any formal labour contract. Importantly, we do not distinguish between different types of shadow contracts: when a job is shadow, it has no legally enforceable duration or termination rules. Legal contract types only become relevant if and when the job is regularised. A shadow job allows firms to avoid paying tax τ and any firing cost. A job that avoids regulation and is not subject to any monitoring generates a value $J^h(\lambda) = (1-\lambda)(y-\omega)$, where the superscript h refers to labour demand in *hidden* jobs, where no tax and contract regulation are enforced, including any form of mandatory firing cost.[20] Yet, shadow jobs are monitored, and we

[20]Shadow jobs are more likely to be subject to "higher" natural turnover than regular jobs, and a clear extension would be to have hidden jobs with an extra turnover term ρ^h, where h refers to shadow jobs.

denote by $J^s(\lambda, \theta^j)$ the value of a λ shadow job, where $\theta^j, j = l, h$ refers to the attitude to go shadow. A shadow job depends on γ, the probability of inspection, and on $C(\lambda)$, the fine imposed on the firm with undeclared work upon inspection. The fine is equal to a fixed term $\bar{c}$ and a term that depends on the duration of the productive relationship λ: $C(\lambda) = \bar{c} + c(\lambda)$. There is a natural assumption that $c(1) = 0$, since the fine on an "instant" job has to be negligible. Furthermore, we assume that $c'(\lambda) < 0$. According to common practice, inspectors will charge a higher fine to firms that appear to be in a longer-lasting employment relationship (and therefore have a job with a lower λ).

The decision to go shadow for a firm of type j with idiosyncratic probability λ is simply

$$J^s(\lambda, \theta^j) = (1-\gamma)J^h(\lambda) + \gamma[J^*(\lambda) - C(\lambda)] + \theta^j > J^*(\lambda), \tag{4}$$

where θ^j is the attitude to go shadow, and $J^*(\lambda)$ is the value function for the optimal labour contract at idiosyncratic duration λ. Equation (4) implies the standard condition that a job goes shadow if the expected tax evasion is larger than the expected cost of evading: $(1-\gamma)(J^h(\lambda) - J^*(\lambda)) + \theta^j > \gamma C(\lambda)$.

To obtain simple analytical insights, we assume that $C(\lambda) = c(1-\lambda)$, so that the cost tends to zero when λ approaches one. The expected surplus from a shadow λ job is $S^s(\lambda, \theta^j) = (1-\gamma)[J^h(\lambda) - J^*(\lambda)] + \theta^j - \gamma c(1-\lambda)$. We can show that–under conditions specified in Proposition 2–the expected surplus is monotone and satisfies the reservation property. Thus, there is a unique value $\lambda^s(\theta^h)$ such that firms with attitude θ^h go shadow if $\lambda > \lambda^s(\theta^h)$.

Proposition 2. *Let the cost parameter* $c > \frac{(1-\gamma)\tau}{\gamma}$*; let the tax evasion attitude* $\theta^h \in \left(0, c - \frac{(1-\gamma)\tau}{\gamma}\right)$*; and let the tax evasion attitude for the low type* $\theta^l < 0$*. It then follows that firms with* $\theta = \theta^l$ *never go "black" shadow, while firms with* $\theta = \theta^h$ *go "black" shadow if* $\lambda > \lambda^s(\theta^h)$*, where* $\lambda^s(\theta^h)$ *solves* $S^s(\lambda^s(\theta^h)) = 0$*, with* $0 < \lambda^s(\theta^h) < 1$*.*

The proof is in Appendix A.2.

The top panel of Figure 3 shows the value function of legal (continuous dark line) and "black" shadow jobs (solid grey line). Legal jobs are characterised by the two reservation productivities λ^{ft} and λ^{v}, thus $J^{*}(\lambda)$ is downward sloping, piece-wise linear and with kinks at λ^{ft} and λ^{v}. For firms with evading attitude, $\theta = \theta^{h}$, "black" shadow jobs (solid grey line) start whenever $\lambda > \lambda^{s}(\theta^{h})$. Formally, the existence of "black" shadow employment implies that firms may open jobs up to the point in which $J^{s}(\lambda^{e,s}(\theta^{h})) = K$, where clearly $\lambda^{e,s}(\theta^{h}) \geq \lambda^{e}$, where the superscript e, s refers to shadow entry. In what follows we define $\lambda^{*,e} = \max[\lambda^{e}, \lambda^{e,s}(\theta^{h})]$ and "black" shadow employment is thus $n^{s} = p[G(\lambda^{*,e}|\theta^{h}) - G(\lambda^{s}(\theta^{h})|\theta^{h})]$.[21]

3.2.1 The Misuse of FWA and Grey Shadow Jobs

Proposition 2 establishes the existence of shadow employment. But when the institutional rules of FWAs allow for instant validation of labour contracts, firms can do even better than passively observe the arrival of a labour inspector. When an inspector arrives, they can immediately sign a voucher for previously undeclared workers. We define these jobs as "grey" shadow.

Since "grey" shadow jobs are shadow as long as the inspector does not show up, let $\tilde{J}^{s}(\lambda, \theta^{j})$ be the value of a "grey" job that has the option to activate the voucher conditional on inspection. Formally, the possibility to go "grey" adds an additional choice, generating an option value conditional on the inspector showing up at the gate. The decision to go "grey" for a firm with idiosyncratic value λ and cost θ is

$$\begin{aligned} \tilde{J}^{s}(\lambda, \theta^{j}) &= (1-\gamma)J^{h}(\lambda) + \theta^{j} + \gamma \left\{ (1-\alpha)(J^{*}(\lambda) - C(\lambda)) + \alpha \max \underbrace{[J^{*}(\lambda) - C(\lambda); J^{v}(\lambda)]}_{\text{option to go "grey"}} \right\} \\ &> J^{*}(\lambda) \end{aligned} \tag{5}$$

With probability α on-the-spot regularisation through vouchers is feasible; conditional on

[21] In Figure A2 we report also the full distribution of Figure 3 to include also the entry margin.

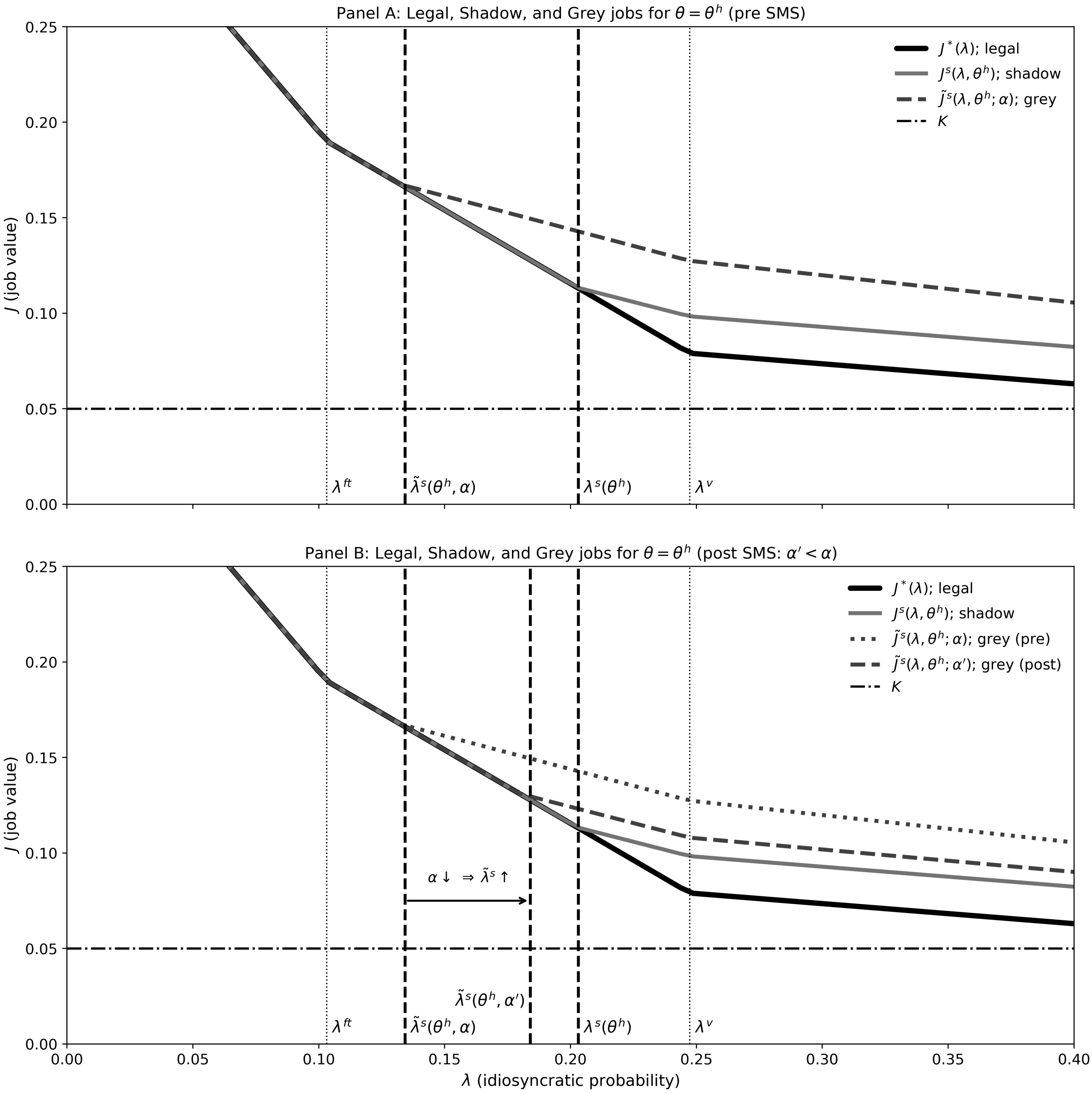


Figure 3: Optimal Contracts and Shadow Employment For Firms with High Tax-evasion Attitude.

Notes: The chart reports the value functions for legal jobs $J^*(\lambda)$ and shadow jobs with misuse $\tilde{J}^s(\lambda, \theta^j)$. Parameters are: $y = 1, \omega = .4, \tau = .25, f = 3$, and $\rho^{ft} = .23, \rho^v = .7, K = 0.05$. The chart considers also the shadow option with linear cost sanction $C = c(1 - \lambda)$ with $c = .92$. $\theta^h = 0.019$ and $\theta^l = -0.081$. The thresholds λ^{ft}, λ^v, λ^s, $\tilde{\lambda}^s$ are indicated in the chart. In the top panel the bold continuous piecewise linear function refers to legal jobs $(J^*(\lambda))$, while the lighter continuous function refers to black shadow jobs for firms with $\theta = \theta^h$. Firms with $\theta = \theta^h$ go shadow for $\lambda > \lambda^s(\theta^h)$. The dotted value function considers also the "grey jobs" and the corresponding reservation $\tilde{\lambda}^s(\theta^h, \alpha^h)$. The bottom panel considers a downward shift from $\alpha = 0.38$ to $\alpha = 0.22$. In moving from the top to the bottom chart the option value of going "grey" is reduced by the change in policy parameter α. See text for function and parameters definitions.

feasibility, the firm chooses whether to regularise or not. When the job is regular, the firm obtains $J^v(\lambda)$ instead of $J^*(\lambda) - C(\lambda)$. The parameter α is a key policy parameter, linked to the technical possibility of declaring and pretending that the job is a voucher. When such a possibility is not available (i.e., when $\alpha = 0$), a grey job is simply a black shadow job. Equation (5) implies that it is impossible during an inspection to pretend that a shadow worker had just been hired with an open-ended or a fixed-term contract; while it is possible to pretend that the worker was hired with a "voucher," and such an option is the economics behind the "grey" shadow jobs. The increase in shadow employment arises from the combination of the short duration of FWAs and the possibility of on the spot activation; tightening administrative requirements (a lower α) reduces "grey" shadow jobs.

For values of λ such that $J^v(\lambda) \geq J^*(\lambda) - C(\lambda)$–that is, when (conditional on feasibility) regularising through a voucher weakly dominates remaining exposed to the fine–the max operator in (5) selects $J^v(\lambda)$, and the condition for misuse becomes:

$$\underbrace{(1-\gamma)\big(J^h(\lambda) - J^*(\lambda)\big)}_{\text{expected tax evaded}} + \theta^j > \gamma \left[\underbrace{(1-\alpha)C(\lambda)}_{\text{expected fine}} + \underbrace{\alpha\big(J^*(\lambda) - J^v(\lambda)\big)}_{\text{loss from voucher regularisation}} \right]. \qquad (6)$$

Comparing equation ((6)) with ((4)) highlights the role of vouchers as an inspection-contingent regularisation technology. When the efficient legal contract for a given job is already a FWA, regularising it through a voucher upon inspection does not entail additional loss relative to $J^*(\lambda)$. For jobs whose efficient legal counterpart is not an FWA, regularisation through a voucher is relatively more costly, as it forces the firm into a shorter expected duration than the efficient legal contract. As a result, there is a threshold $\tilde{\lambda}^s(\theta^h)$ above which firms optimally regularise shadow work through vouchers when inspected.

For any $\alpha > 0$, it is obvious that $\tilde{\lambda}^s(\theta^h) < \lambda^s(\theta^h)$, which means that the possibility to validate vouchers on the spot increases the amount of total shadow employment, generating grey jobs. In the top panel of Figure 3 grey shadow jobs are indicated by the dotted grey line, and it is clear that there is more shadow employment when misuse is feasible. Moreover, the behaviour of the firm during an inspection is indicative of their type θ^j: only θ^h firms misuse vouchers.

While stronger enforcement ($\gamma \uparrow$) reduces incentives to keep jobs in shadow, this deterrent effect is weaker when on-the-spot grey regularisation is easier ($\alpha \uparrow$). In our setup, the interaction is governed by

$$\frac{\partial^2 S}{\partial \alpha \, \partial \gamma} = C(\lambda) - \big(J^*(\lambda) - J^v(\lambda)\big),$$

so, if the expected fine exceeds the voucher-regularisation loss, $C(\lambda) > J^*(\lambda) - J^v(\lambda)$, then increases in α attenuate the impact of γ on shadow work (this is always true when $J^*(\lambda) = J^v(\lambda)$). Intuitively, when firms can quickly relabel detected shadow work as vouchers, inspection risk carries less effective punishment. A reduction in deterrence reduces the gap in shadow work between high and low enforcement places.

3.2.2 Testable Predictions

Thus, we have the first key testable prediction.

> **Testable Prediction 1 (Pre-SMS Inspector Arrival and Grey-Job Vouchers):** When inspectors arrive at the gate, firms with $\theta = \theta^h$ exhibit an immediate increase in the use of vouchers linked to "grey" jobs, whereas firms with $\theta = \theta^l$ do not.

Note that we can also quantify the misuse of vouchers. With the option to misuse we have $\tilde{J}^s(\tilde{\lambda}^{e,s}(\theta^h)) = K$, where clearly $\tilde{\lambda}^{e,s}(\theta^h) \geq \lambda^{e,s}(\theta^h)$. In what follows we define the entry margin with grey jobs as $\hat{\lambda}^{e,s} = \max[\lambda^e, \lambda^{e,s}(\theta^h), \tilde{\lambda}^{e,s}(\theta^h)]$ and misused employment is thus

$\tilde{n}^s = p[G(\tilde{\lambda}^{e,s}|\theta^h) - G(\tilde{\lambda}^s(\theta^h)|\theta^h)]$. Further, the change in the likelihood of using a voucher when the inspector shows up is instead:

$$\Delta\tilde{q}^v(\theta) = \begin{cases} G(\tilde{\lambda}^{e,s}(\theta^h,\alpha)|\theta^h) - G(\tilde{\lambda}^s(\theta^h,\alpha)|\theta^h) > 0 & \text{if } \theta = \theta^h, \\ 0 & \text{if } \theta = \theta^l\ . \end{cases} \tag{7}$$

Thus, the average jump when the inspector shows up is the product of the share of low tax morale firms p and the likelihood that the job falls in the range when the firm uses grey shadow jobs:

$$\beta = p\Delta\tilde{q}^v(\theta^h). \tag{8}$$

Without an additional moment from post-abolition differential responses between misusing and compliant firms, one cannot separately identify p and $\Delta\tilde{q}^v$ from equation (8) alone. In the empirical analysis, we therefore use an additional moment around the abolition of vouchers to discipline this decomposition and try to separate the extensive margin (p) from the intensive margin ($\Delta\tilde{q}^v$).

Next, we consider what happens to the misuse of vouchers when the policy setting changes, and it becomes more difficult to activate a voucher on the spot. In the model, this is equivalent to a reduction in the parameter α. The key comparative static is thus the derivative $\frac{\partial\tilde{\lambda}^s(\theta^h)}{\partial\alpha} < 0$. This implies that as α falls, the grey-shadow threshold moves to the right (see Panel B of Figure 3), reducing spot validation and, potentially, the amount of grey and thus shadow jobs.

> **Testable Prediction 2 (Post-SMS Inspector Arrival and Grey-Job Vouchers):** With the introduction of the text message requirement (a fall in α), upon inspection, firms with $\theta = \theta^h$ reduce their spot validation of vouchers. Firms with $\theta = \theta^l$ do not.

The reduction in the likelihood of misuse comes from a simple comparative static result: $\frac{\partial G(\tilde{\lambda}^s)}{\partial \lambda^s}\frac{\partial \tilde{\lambda}^s(\bar{\theta})}{\partial \alpha} < 0$. Furthermore, the corresponding reduced change in the likelihood of using a voucher when the inspector shows up is again given by equation (8) but with a reduced $\alpha' < \alpha$. Also this event can be taken to the data. The bottom panel of Figure 3 shows how the value of jobs varies when α decreases, thus when the legislation required that employers inform the administration one hour in advance about using a voucher. As mentioned earlier, even if spot validation decreases, firms keep the option to disguise shadow work by buying a voucher per day per worker. In this case, we would expect to observe i) an increase in the use of vouchers when the SMS requirement was introduced and ii) little change in legal jobs.

The last exercise we perform considers what happens when vouchers are abolished. When vouchers are banned, firms can no longer use the most flexible legal margin (margins λ^v and $\tilde{\lambda}^s(\theta^h)$ disappear). So both firm types reallocate some jobs toward the next-best legal flexible contract: fixed-term. In the absence of vouchers for θ^l firms, the fixed-term-job margin is implicitly defined by $J(\lambda^{ft,no-v}) = K$. Analytically, we have that $\lambda^{ft,no-v} > \lambda^v$, so that θ^l firms expand fixed-term contracts. For θ^h, previously misused vouchers are transformed into fixed-term contracts, all the way until shadow work becomes preferable: the right margin for fixed-term jobs moves from $\tilde{\lambda}^s(\theta^h, \alpha')$ to $\lambda^s(\theta^h)$.

This result implies an average increase in fixed-term employment for firms with $\theta = \theta^h$ equal to $\Delta n^{ft}(\theta^h) = n[G(\lambda^s(\theta^h)|\theta^h) - G(\lambda^{s'}(\theta^h, \alpha')|\theta^h)] > 0$ while the increase in fixed-term employment for firms with $\theta = \theta^l$ is $\Delta n^{ft}(\theta^l) = n[G(\lambda^{ft,no-v}|\theta^l) - G(\lambda^v|\theta^l)] > 0$. In other words, fixed-term employment increases for both types of firms.

Testable Prediction 3 (Voucher Ban): When vouchers are abolished, fixed-term employment rises for both firm types. The magnitude of the increase depends on the cutoff $\tilde{\lambda}^s$ and on the distributions $G(\lambda \mid \theta^h)$ versus $G(\lambda \mid \theta^l)$. If some fixed-term jobs are "grey" and $G(\lambda \mid \theta^h)$ first-order stochastically dominates $G(\lambda \mid \theta^l)$ over the relevant range (i.e., low-tax-morale firms with $\theta = \theta^h$ have higher-λ jobs), then low-tax-morale firms exhibit a larger increase in fixed-term employment.

If low-tax-morale firms have relatively more high-λ jobs in the relevant region (shorter/less stable jobs), they relied more on vouchers before the ban, so they show a larger post-ban increase in fixed-term jobs.

Heterogeneity in firms' responses can arise from differences in enforcement intensity (γ), in tax morale (θ), or in the distribution of job durations $G(\lambda)$. We already mentioned that γ becomes less important when vouchers can be misused. But a higher prevalence of short-duration jobs is predicted to generate greater misuse of vouchers when the inspector shows up.

The reduction in shadow when vouchers are banned is given by $\Delta\tilde{n}^s = [G(\lambda^s(\theta^h)|\theta^h) - G(\tilde{\lambda}^s(\theta^h, \alpha')|\theta^h)] + [G(\tilde{\lambda}^{*,e}) - G(\lambda^{*,e})]$ where the first part refers to the transformation of grey jobs into regular jobs, while the second part is a reduction in entry due to the impossibility of creating grey jobs.

While total shadow work is predicted to decrease, in the absence of the option to hide it with grey contracts, the detected shadow work may actually increase.

In the rest of the paper, we test these empirical predictions.

4 Data

We use data from the INPS archives, including firm-level employment data, daily firm-level voucher usage, and labour inspection records. Our model has clear predictions about the use of vouchers when firms receive a labour inspection. This is why the analysis focuses on firms that have been inspected at least once and that have used at least one voucher during our sample period.[22]

The sample construction follows these steps. We start with the universe of firm-level employment data for the years 2014 to 2017. As shown by the timeline of the regulatory changes to the FWA (Figure 1) and the time series of vouchers used (Figure 2), following almost complete liberalisation, the use of vouchers peaked right before their abolition. Later, we merge this firm-level employment data with both the universe of labour inspections and the universe of vouchers used by firms.[23]

The INPS institute performs labour inspections to detect full or partial evasion of social security contributions. We have information on the day the inspection started and the outcome of the inspection (whether a fine was levied and its amount).

For each firm, we also know how many vouchers have been used each day. In order to maximise the time frequency of our data, we construct firms' daily use of vouchers around labour inspections. Later, when we analyse firm-level employment outcomes, the finest level of aggregation we can use is at the firm and month level.

The summary statistics in Table 1 are useful for describing how labour inspections and the use of vouchers determine the type of firm that is later analysed. Each row represents monthly averages between January 2015 and September 2016 (the period before the SMS requirement reform) of various firm-level employment measures -both regular work and vouchers- and each

[22] Firms that never use vouchers may require a more stable workforce (a lower λ in our model) and would not generate any empirical variation in the use of vouchers.

[23] These two sources of data are not available within the VisitInps program and have been directly managed by two of the authors who worked for INPS, Edoardo di Porto and Paolo Naticchioni.

Table 1: Summary Statistics

Firms:	All	Vouchers	Inspected	Both	Misusing	Compliant
Worforce	7.171	7.212	43.511	40.818	42.737	39.411
Monthly Wage bill	14,329	11,968	93,170	63,042	61,664	64,051
Average wage	1,346	1,271	1,316	1,247	1,231	1,258
Monthly Wage Bill (FTE)	15,894	13,737	102,458	75,886	78,431	74,021
Workforce (FTE)	6.3	6.1	38.5	33.7	33.7	33.7
Temporary workers	1.1	1.8	9.4	16.5	14.4	18.0
Part-time workers	2.1	2.7	11.8	17.0	20.8	14.1
Permanent workers	6.1	5.4	34.1	24.3	28.3	21.4
Full-time workers	5.0	4.5	31.7	23.9	21.9	25.3
Monthly number of vouchers	27.9	97.9	306.8	252.9	271.3	239.4
Voucher per workforce (FTE)	4.461	15.928	7.979	7.500	8.050	7.097
Number of firms	1,840,339	416,942	40,121	3,472	1,469	2,003
Fraction	100%	22.66%	2.18%	0.19%	0.08%	0.11%

Notes: This table shows averages for different samples using 2015 and September 2016 data. The samples are the following: "Full" is the universe of private firms in 2015 and 2016; "Vouchers" are firms that have used at least one voucher in those years; "Inspected" are firms that have been inspected between January 2015 and September 2016 (almost all are inspected only once); "Both" are firms that use at least one voucher and have been inspected. Presumably "Misusing" and "Compliant" firms either increase or decrease the use of vouchers upon inspection. "FTE" stands for full-time-equivalent.

column represents a different sample.

In Column 1, we start with the universe of 1.84 million firms. The average firm in the INPS archives employs 7.2 workers and uses 28 vouchers per month. If each voucher covered one hour of work, the vouchers would represent only a small share of labour; however, we provide evidence consistent with vouchers being used to conceal additional undeclared work hours.[24] Most workers have permanent and full-time contracts.

About 23% of firms (a little more than 400,000) use at least one voucher (Column 2, Vouchers). For firms using vouchers, the average size is also close to 7, but with more temporary and part-time workers and fewer permanent and full-time ones. These firms use an average of 98 vouchers, or roughly 16 vouchers per full-time equivalent worker.

The third column (Inspected) refers to inspected firms, which, due to the small probability of inspection, represents only 2.18% of firms.[25] Since larger firms are more likely to be inspected, the average workforce increases to 43.5 workers. In addition, inspected firms use a large number of vouchers, on average 307 vouchers per month, or 8 vouchers per full-time equivalent worker.

In order to analyse how the use of vouchers evolves around the onset of labour inspections, we focus on firms that have been inspected and use at least one voucher over the entire period. We are left with 3,472 firms, or 0.19% of firms (column "Both"). The number of employed workers is similar to that of the inspected firms, but with a higher share of part-time contracts and temporary contracts, and an average monthly number of vouchers of 252.9 (7.5 per full-time equivalent worker). This sample allows us to analyse how voucher–using firms – which represent about one quarter of Italian firms – respond to an inspection.

In the last two columns, we divide the 3,472 firms into firms that appear to use grey jobs

[24] Unfortunately, the available data do not allow us to estimate how many hours of work tend to be under reported due to the availability of flexible work schedules.

[25] 60% of inspections resulted in fines and the average sanction was around €20,000. Later, we are going to analyse how such fines vary depending on whether firms use vouchers.

and thus increase voucher use at the time of inspection ("Misusing") and those who did not ("Compliant"). The division depends on the firm's use of vouchers when a labour inspector shows up, something we are going to explain later.

5 Empirical Strategy

To keep the empirical evidence tightly linked to the model, we organise this section around Testable Predictions 1–3. Prediction 1 concerns the pre-SMS inspection-day jump in voucher use; Prediction 2 concerns the disappearance of that discontinuity after SMS; Prediction 3 concerns post-abolition substitution toward an increase in regular flexible contracts and higher fines in case of inspections.

To test these predictions the empirical strategy proceeds in several steps, exploiting randomness of labour inspections and some policy reforms such as the introduction of the SMS notice and the abolition of vouchers.

At first, to test prediction 1 we exploit the random timing of labour inspections (Section 5.1). Second, we discuss the empirical specification of the event study (Section 5.2) and, third, we explain how those estimates are used to identify misusing firms and introduce the difference-in-difference models that compare misusing against compliant firms (Section 5.3), comparison that will be used to test predictions 2-3. Section 6 presents the results, examining whether firms, exploiting the randomness of labour inspections (Section 6.1), increase voucher use when an inspection starts (Section 6.2), and what happens to different types of regular and irregular work when the SMS is introduced (Section 6.3) and when vouchers are abolished (Sections 6.4 and 6.5).

5.1 Random Timing of Inspections

We are going to see that inspections are rare events, and their timing is intended to be unpredictable, allowing for a comparison of firms' behaviour before and after inspections.

A randomly selected firm has a 1 in 130 chance of being inspected in a given year and about a 1 in 50,000 chance of an inspection starting on a given day. Although some firms are more likely to be inspected than others, for example, larger firms, to be effective, the timing of inspections has to be unpredictable. Thus, from a firm's perspective, the first day that labour inspectors enter the firm's premises is as good as random.

This assumption is testable. If the timing of inspections is as good as random, the vector of observable characteristics (X_j) of a firm j should be unable to predict the exact date (t_j) of a labour inspection. We use a balance test, the joint F-test, to determine whether all characteristics in the following cross-sectional regression have no predictive power $(\beta = 0)$: $t_j = \alpha + \beta' X_j + \epsilon_j$.

Given the random timing of inspections (see Section 6.1), our equation (8) in the model suggests a fairly simple test to determine whether an inspected firm is using vouchers to hide undeclared work: when an inspection starts, firms employing undeclared workers should increase voucher use in order to hedge the risk of sanctions. We compare the daily use of vouchers just before and after an inspection (the "treatment"), following the firm's behaviour before and after the inspection.

5.2 Random Timing and Event Studies

Once we have established that inspections happen at random times, in the second part, we are going to set up the event study to test the model predictions about Pre and Post-SMS use of vouchers when the labour inspector arrives (Testable Predictions 1 and 2).

We analyse the vouchers used by the firm j between 180 days before and 90 days after an inspection that occurs on day t.[26] We split the event studies into two periods, before and after October 2016, which is when firms had to inform the Social Security Administration an hour in advance before using a voucher.

Since a single voucher is sufficient to avoid sanctions, our outcome variable is equal to one when on a given event day τ the firm j uses at least one voucher, and 0 otherwise ($DV_{j,\tau} = 1\{\#Vouchers_{j,\tau} > 0\}$).[27] Of the 180 days before the inspection date, the first 90 days will serve as the control period:[28]

$$DV_{j,\tau} = \sum_{k=-90}^{90} \beta_k D_{\tau+k} + f(t) + \epsilon_{j,\tau} \ . \tag{9}$$

$D_{\tau+k}$ is a dummy variable equal to one for event day $\tau + k$ and zero otherwise. Additionally, since the time series of FWAs is far from stationary (see Figure 2), it is necessary to control for the calendar time $f(t)$. We start using several calendar-time fixed effects (e.g. year, month, and day of the week) and show that the results do not differ when using calendar-day fixed effects.[29]

The unpredictability of the timing of inspections is crucial in setting the correct specification for our model. This is because focusing on just treated firms that are treated at different times, conditional on firm fixed effects, calendar time, and event time of the inspection, is collinear. Fortunately, when the timing of the event is random (see Section 6.1) the timing of treatment is orthogonal to the characteristics of the firms and to their time-invariant inter-

[26] As mentioned, the last phase of labour inspections can approximately last up to 90 days; moreover, labour judges have deemed that 90 days represent the reasonable duration of labour inspections (Del Vecchio, 2019).

[27] Here $1\{A\}$ is a dichotomous variable equal to 1 when statement A is true, and 0 otherwise. Results are similar when using the daily number of vouchers as the outcome, although the effect is primarily driven by the extensive margin. This is consistent with the presence of a limited number of undeclared workers per (declared) firm.

[28] Given that the days event-time effects leading to the inspections are precisely estimated to be zero, a longer baseline period is going to lead to more precisely estimated coefficients. But shorter baseline periods lead to very similar results.

[29] We also document how the results differ when we do not control for time or control for a simple linear time trend.

cept (see Borusyak et al., 2024). Our strategy is similar to the identification used in Parker et al. (2013), which exploits the random timing of the disbursement of the 2008 Economic Stimulus Payments in the United States.

It is also worth noting that, due to such randomness, we do not have to specify a two-way fixed effects model; thus, we do not have to worry about the well-known possible biases arising from this design (see Goodman-Bacon, 2021, Sun and Abraham, 2020). When, as a robustness check, we do specify a two-way fixed effects model, it is only parametrically identified. A relatively long baseline period (90 days) combined with the non-linear function of event time (for example, event-time dummy variables) breaks the perfect collinearity between event-time and calendar time, but the fix is not perfect. Such near-collinearity is measurable. Without firm fixed effects, time effects explain only 4% of the variation in the post-inspection variable (our main explanatory variable), but when we add firm fixed effects the R^2 goes up to 55%. In other words, because of near-collinearity, using both firm effects and calendar-time effects captures most of the random variation in the timing of inspection.

We are going to discuss later how we deal with these potential biases. Importantly, with random timing of inspections, we can use the main result from Athey and Imbens (2022), which shows that the standard Difference-In-Differences estimator is an unbiased estimator of a weighted average of different causal effects, including the effect of changing from never being inspected to being inspected in the first period, or changing from being inspected later to being inspected earlier in time. Since in our setup there are no reasons to believe these effects differ, apart from the collinearity issues, a simple Difference-In-Differences identifies the average causal effect.[30]

Moreover, under the design-based assumption that each unit's adoption (or inspection) date is randomly assigned, the simple before-after difference in means remains an unbiased estimator of a well-defined average treatment effect on the treated (the inspected firms).

[30] Athey and Imbens (2022) also show that the difference-in-difference standard errors are conservative.

The data contain information about the firms, which allows us to look for heterogeneous effects. Given our simple test for the misuse of vouchers, the heterogeneity analysis will inform us whether the misuse of vouchers is concentrated in certain parts of the country, certain sectors, and specific types of firms (large vs. small, with more or less use of temporary contracts). In order to do this, we simply test whether pre-SMS effects differ between economic sectors and firm characteristics. Given the apparently static treatment effects and the lack of pre-trends, for the heterogeneity analysis, we collapse the post coefficients over time and use the whole pre-inspection period as baseline:

$$DV_{j,\tau} = \beta D_{\tau \geq 0} + f(t) + \epsilon_{j,\tau} \ . \tag{10}$$

5.3 Misusing and Compliant Firms

In the third part, we use the predictions of the model to identify the misusing firms: those that increase their use of vouchers after the inspection. This allows us to test the model's predictions about the differential behaviour of θ^l and θ^h firms.

In order to spot them, we exploit the fact that treatment effects appear to be stable over time. With firm-specific constant treatment effects, we can model the change in the probability of using a voucher to be constant over time but heterogeneous across firms:

$$DV_{j,\tau} = \beta_j D_{\tau \geq 0} + f(t) + \epsilon_{j,\tau} \ , \tag{11}$$

where β_j represents the firm-specific post-pre inspection differences in the likelihood of using vouchers.

Based on the estimated $\hat{\beta}_j$s, we identify misusing firms and highlight how they respond to the introduction of the SMS[31] requirement and the abolition of vouchers. In particular, we

[31] SMS are Short Messages or Text Messages in mobile telecommunication.

combine the results from the previous model with the October 2016 SMS requirement and the March 2017 abolition of vouchers. We define misusing firms as those that, following an inspection, increase their probability of using a voucher by at least η: $\widehat{M}_j^{\eta} = 1\{\hat{\beta}_j > \eta\}$. In the baseline analysis we use $\eta = 0$, and later we test whether the results are robust to the choice of more stringent cut-offs, $\eta > 0$.

In other words, we use behavioural changes driven by the inspections to identify misusing or low-tax morale firms: for each inspected firm, we compute its average use of vouchers before and after the inspections and classify firms into those that, on average, increase their use and those that do not.[32] Our definition of misusing firms is subject to misclassification, both Type I and Type II, which biases the estimates towards 0. Later, we will take advantage of the possibility of changing our definition of misusing firms.

The labour demand model has predictions about how misusing firms ($M_j = 1$) react when the SMS requirement is introduced (Testable Prediction 2) and when vouchers are abolished (Testable Prediction 3). In this latter case, firms might: fall back to signing legal contracts (mainly the next most flexible contracts with respect to vouchers, such as the temporary contracts); revert to hiding the entire work relationship; abandon the low-productivity jobs.

We use simple difference-in-differences, before and after October 2016 (SMS requirement) or March 2017 (abolition of vouchers) between firms that presumably misused FWAs and those that did not.[33] Since all firms share the same event date, we are not in a staggered design and do not have to worry about biases due to dynamic treatment effects (see Sun and Abraham, 2020). To assess the parallel trend assumption, we estimate event study differences with leads and lags. The number of lags is limited by the period spanned by the data, and we exclude the event time $\tau = -2$ (respectively, August 2016 and January 2017,

[32] While firms that do not use vouchers may in principle be used as non-misusing firms, Table 1 shows that they are quite different in terms of observables and possibly unobservables from those firms that use vouchers.

[33] We are implicitly assuming that misbehaviour is time-invariant, and any deviation from this assumption would lead to downward biased estimates.

allowing for some anticipation effect).

The outcomes available at the monthly level (m) are i) the total number of vouchers used, ii) the log of the total wage bill (including vouchers) and the total number of workers with the following contracts: iii) part-time, iv) full-time, v) fixed-term, vi) fixed-term and part-time, and vii) open-ended. The event study difference-in-difference controls for firm and year by month fixed effects:

$$Y_{j,t}^{m} = \sum_{k \neq -2} \pi_k \widehat{M}_j^0 \times D_{\tau(t)+k} + \mu_j + \mu_t + \varepsilon_{j,t} \ , \tag{12}$$

where τ is the event period, which is 0 either in October 2016 or in March 2017, and $D_{\tau+k} = 1$ in event period $\tau + k$ and 0 otherwise.

To better understand the issue of misclassification in misusing firms $\widehat{M}_j$, we use a constant difference-in-difference model:

$$Y_{j,t}^{m} = \pi \widehat{M}_j^{\eta} \times D_t + \mu_j + \mu_t + \varepsilon_{j,t} \ , \tag{13}$$

where D_t is a indicator that equals one after the abolition of vouchers. Defining $p(\eta)$ and $q(\eta)$ to be type I and II errors of the misclassified and unobserved variable M_j, it can be shown that, under constant treatment effects, $\frac{\widehat{\pi}(\eta)}{1-p(\eta)-q(\eta)}$ converges in probability to π. This implies that, with constant treatment effects, the η^* that maximises $\widehat{\pi}(\eta)$ minimises misclassification bias. Later, we discuss whether η can tell us something about the fraction of firms that misuse FWAs.

In our final analysis, we look at how the introduction of the SMS requirement, as well as the abolition of vouchers, has influenced under-reporting of work. Since shadow black jobs are not directly observable, we need to rely on labour inspections. Yet, only a handful of firms are inspected more than once in our data, therefore we cannot compare misusing with compliant firms, but need to rely on a coarser definition of treatment.[34]

In particular, we compare firms based on whether they have ever used a voucher before the SMS requirement, and thus had at least the option to misuse. An observation is defined as a labour inspection between 1 January 2016 and 31 December 2017, which produces 20,819 observations. The treated firms are those that have used vouchers in the pre-SMS period (4,269 observations). We use the evaded contribution as an outcome variable to measure under-reporting and define three main treatment periods: Pre-SMS requirement (1 January to 16 October 2016, which will be the omitted time period), post-SMS requirement (17 October 2016 to 17 March 2017), and post-abolition (18 March 2017 to 31 December 2017). Consistently with Testable Prediction 3, we expect fines paid upon inspections after vouchers' abolition to be higher than in the baseline period.

6 Empirical Evidence

6.1 Random Timing of labour Inspections

To test the prediction 1 of the theoretical model we start by testing whether the timing of the labour inspection is predictable. Appendix Figure A3 shows the distribution of the day of the year inspections take place. Inspections occur throughout the year and are fairly uniformly distributed, although they are slightly more likely to occur during the summer.

The panels in Figure 4 show the linear regression coefficients of different firm characteristics on the day of inspection. In the left panel, we only control for year fixed effects, to allow labour inspectors to have different targets in different years. Only three variables are

[34]Based on Table 1, the number of the inspected firms is 40,121, or 2.18% of the population of firms. 2,955 firms are inspected more than once over the same period, representing 0.16% of firms. Notice that if inspections were random events the likelihood to be inspected twice would be 2.18% to the power two, or 0.048%. This implies that firms that are inspected once are more likely to be inspected again, and this might explain persistency in the effects of inspections.

significantly different from zero, and each characteristic predicts at most a difference of a few days on the date of inspection. The largest difference is in the construction sector, where inspections tend to happen two weeks earlier. Adding a semester dummy is sufficient to drive all these differences towards zero even further. Within 180 days, all differences are, with the exception of the residual sector "Other," smaller than 3 days. In other words, the exact day that inspections take place appears to be unpredictable.

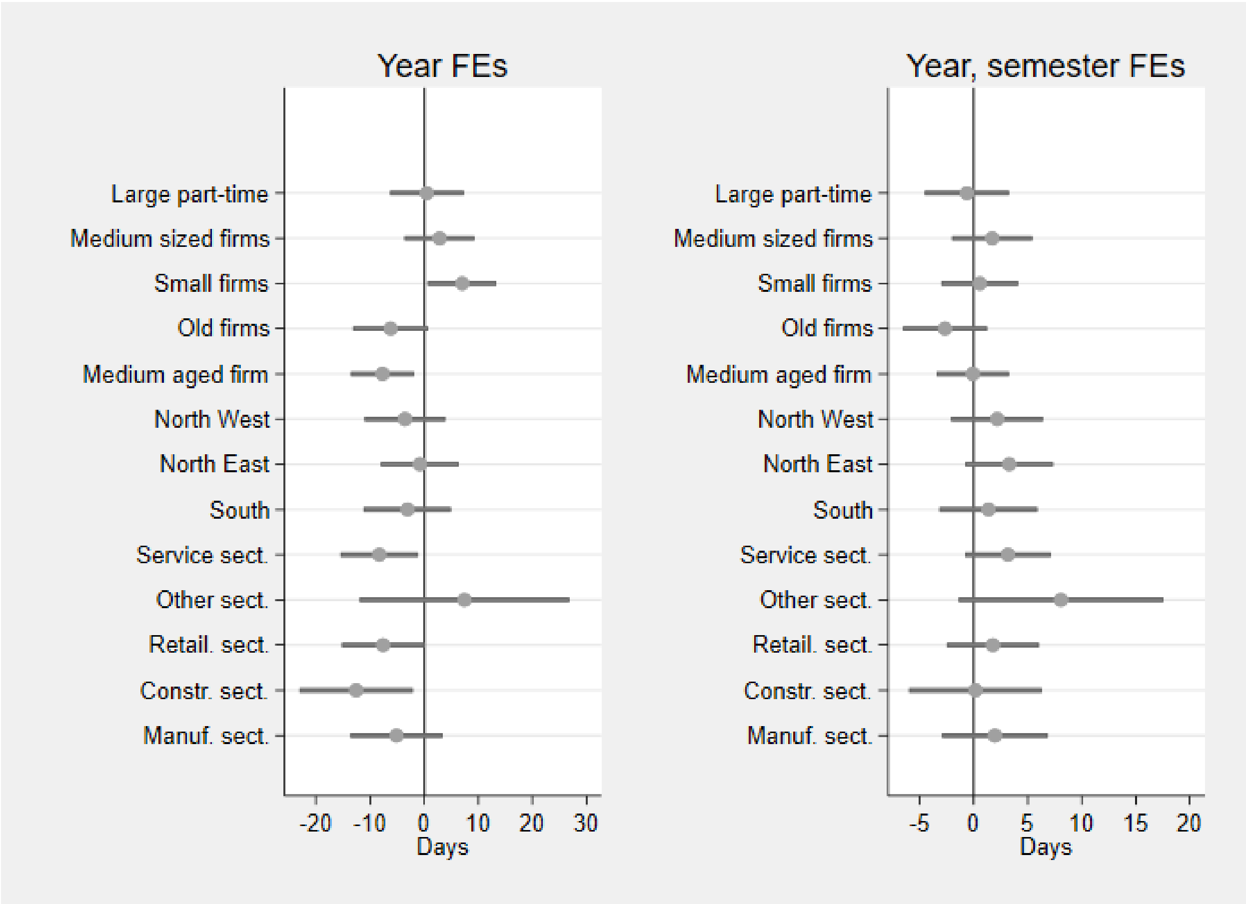


Figure 4: Balance Test

Notes: The figure plots the coefficients of day of inspection on firm characteristics. The regression results shown on the left control for year fixed effect, the ones in the second add also semester fixed effects. The vertical bars represent the 95% confidence intervals based on robust standard errors. The p-values for the F-test that sets all coefficients equal to zero are respectively 0 and 72%.

6.2 The Misuse of FWAs: Evidence from Labour Inspections

Given random timing, we can estimate differences in the probability of using vouchers around the inspection time. Figure 5 plots the event-study differences from Equation (9) using a linear probability model. Each point represents $\widehat{\beta}_{\tau+k}$, that is, the difference in voucher use between event date $\tau+k$ and the days between 90 and 180 before the inspection (the excluded event times). Upon inspection, there is a clear change in the likelihood of using vouchers. Moreover, the evidence suggests that conditional on year, month, and day of the week fixed effects, there are i) no pre-trends in the use of vouchers prior to the inspection, ii) no major anticipation effects, and iii) fairly stable and persistent treatment effects, consistently with Testable Prediction 1. Once inspectors arrive, the evidence suggests that a misusing firm with undeclared workers needs to justify the continued presence of those workers not just that day but for the subsequent days they remain on site. In that sense, the firm's optimal response is not a one-day "insurance" action, but a sequence of voucher activations that can last as long as the workers remain and/or the inspection risk remains salient.

With respect to point i), Appendix Figure A4 shows how important it is to control for time effects. Without such controls, the event dummies will capture part of the increasing trend in the use of vouchers (right panel). Yet, the left panel shows that adding a simple linear time trend is enough to centre the pre-period around zero (left panel).

The same is true for point ii): without time controls, the trends generate what looks like an anticipation effect. With respect to point iii) the increase in the likelihood of using vouchers immediately after an inspection is 0.88 percentage points (SE 0.16), which corresponds to a relative increase of approximately 18%. The greatest changes occur on the day of inspection and the day after, respectively, 1.5 (30%) and 1.4 percentage points (29%). If we consider that once the inspection has started, firms may also have the option to ask undeclared workers to stay home, these findings correspond to large effects. There is possibly a very

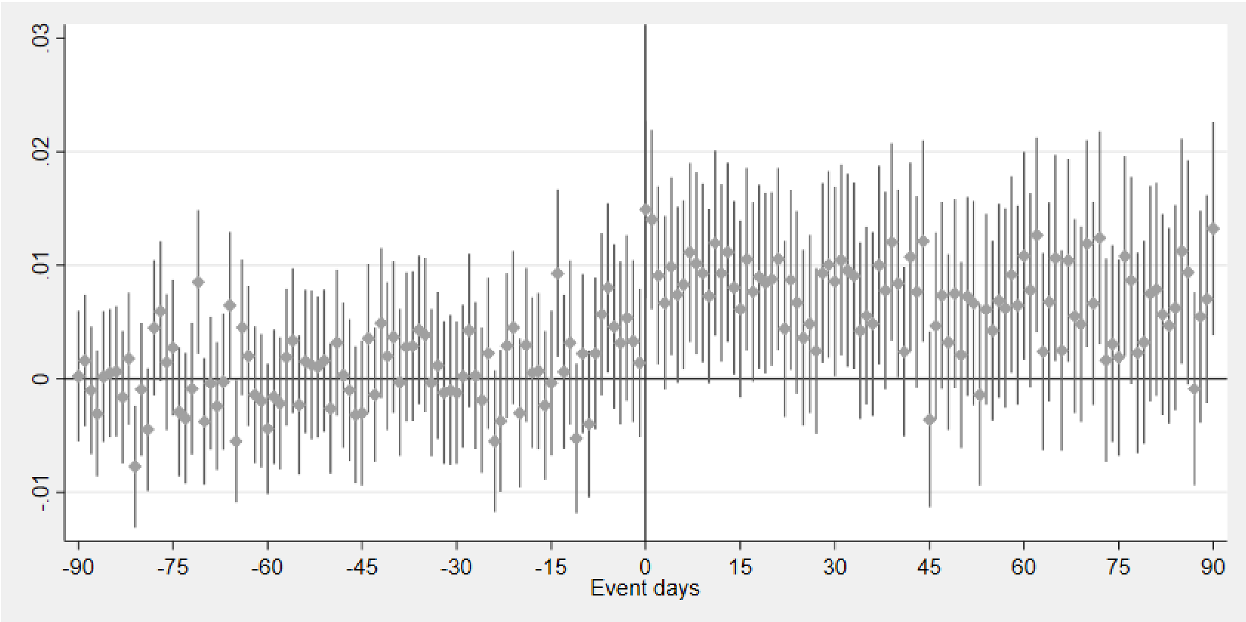


Figure 5: Event Study pre-SMS

Notes: The figure plots event study coefficients, where the event is a labour inspection. The excluded time period is between 180 and 90 days prior to the inspection. The regression controls for year, month, and day of the week fixed effects. Standard errors are clustered at the firm level.

slight downward trend in the effects. For the first three weeks, almost all daily effects are significantly different from zero, while there is more noise after the first 45 days. There are still several significant effects after that event date; interestingly, there is an additional spike at 90 days, which, as mentioned earlier, sometimes corresponds to the prescribed end of the last phase of inspections.[35]

This inspection-day discontinuity is the empirical signature of inspection-contingent misuse: when on-the-spot validation is feasible, a subset of firms activates vouchers to regularise otherwise undeclared work. We refer to this subset as "misusing" firms.

Having many pre-inspection periods allows us to perform randomisation inference. Focusing on pre-inspection data ($\tau < 0$), we sequentially generate fictitious inspection dates for $-120 \leq \tau \leq -30$ and estimate

$$DV_{j,\tau} = \beta_k D_{\tau-k\geq 0} + f(t) + \epsilon_{j,\tau} \ , \tag{14}$$

where $D_{\tau-k\geq 0} = 1$ when $\tau - k \geq 0$ and zero otherwise. The histogram of all placebo $\widehat{\beta}_k$'s, shown in Appendix Figure A5, is centered around zero and is far from the vertical line, which corresponds to the estimated $\widehat{\beta}_0 = 0.88$. Sampling noise is unlikely to have generated such a large change in behaviour.

The data contain information about the firms, which allows us to look for heterogeneous effects. In particular, we test whether pre-SMS effects differ across economic sectors and firm characteristics. Given the static treatment effects and the lack of pre-trends, we collapse the treatment effects and use the whole pre-inspection period as a baseline:

$$DV_{j,\tau} = \beta D_{\tau\geq 0} + f(t) + \epsilon_{j,\tau} \ . \tag{15}$$

[35] We know that inspections should not last more than 90 days since the preliminary report. While we do not have information on the duration of the preliminary phase, whenever such phase takes less than a day the maximal duration corresponds to our event date 90. This fuzziness in the prescribed end date (the actual date would be endogenous) makes it hard to empirically exploit the end of inspections.

Table 2 shows that the increase is about the same in the retail sector, the tourism sector, and the manufacturing sector. For the "Other sectors" the jump is slightly lower, while it is completely absent in the construction sector. It turns out that the construction sector is highly dependent on procurement jobs and subcontracting, and since 2015, with "the Jobs Act," vouchers cannot be used within such contracts. The aim was to reduce the number of job accidents.[36] This is likely to explain why, in the construction sector, firms have the lowest probability of using a voucher (0.02 per day).

Table 2: Vouchers and labour Inspections by Sector

Sector	(1) Manufacturing	(2) Construction	(3) Retail	(4) Tourism	(5) Other Services
Post-Inspection	0.013***	-0.002	0.012**	0.011***	0.006*
	(0.003)	(0.003)	(0.006)	(0.002)	(0.003)
Constant	0.032***	0.021***	0.030***	0.054***	0.051***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.004)
Observations	157,329	98,087	208,194	614,758	255,718
R-squared	0.014	0.013	0.010	0.022	0.016
Mean dep. var.	0.0381	0.0201	0.0352	0.0592	0.0541

Notes: Linear probability model of using at least one voucher, daily data with year, month and day of the week fixed effects. Clustered standard errors (by firm) in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

In Appendix Table A1 we perform additional heterogeneity regressions. Columns 1 to 4 show that the jump is relatively constant across Italian regions, although it is slightly larger in the more productive North than in the South, with the Centre of Italy being in between. Thus, while the South is usually perceived to be more prone to tax evasion, it does not misuse vouchers more than the other parts of Italy; rather, slightly less. Consistent with the model, when on-the-spot misuse is feasible (higher α), the deterrent effect of enforcement differences (γ) is attenuated, so cross-region differences in jump magnitudes may remain

[36] According to the Italian Work Injury Institute INAIL ("Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro)" the share of deadly accidents in the construction sector is the largest. For example, in 2019 around 23% of deadly accident occurred in the construction sector, a sector that employs only 6.6% of the workforce.

limited even when enforcement intensity differs.

Columns 5 to 7 show that medium-aged firms (those that started between 5 and 14 years earlier) are more likely to use vouchers to cover undeclared work compared to young and old ones. Firm size is highly predictive of the size of the effects, with large firms (more than 15 employees) being the ones with the largest jumps (column 10). Column 11 shows that the jump is about 40% larger for firms whose share of part-time workers is above the median. This is in line with the opinion of many labour inspectors that part-time work is sometimes used to hide what are truly full-time workers, as it lowers the social security contributions and the tax burden. Finally, according to the law, firms' use of temporary workers should not exceed 20% of the open-ended ones. Yet, in our sample (but this is true more generally) more than 50% of firms are above such limit, suggesting that enforcement of this regulation is limited. Under the assumption that such behaviour is indicative of lawless behaviour, Columns 12 and 13 show, indeed, that such firms above the threshold of 20% tend to have a more pronounced jump (1.1 p.p. against 0.85 p.p.).

In the last robustness check, we add firm fixed effects that capture part of the unobserved heterogeneity in the use of vouchers. As mentioned earlier, with random timing, adding firm fixed effects is unnecessary and may do more harm than good. This can be seen in Appendix Figure A6, which shows that firm fixed effects generate pre-trends in the use of vouchers. Since firm fixed effects introduce collinearity between event time and calendar time, it appears as if some of the time effects are now captured by the event-time effects, generating a rotation in all the differences. The bias due to the rotation increases as we move further away from the date of inspection, reducing all the post-inspection changes.

One way to reduce collinearity bias is to focus on differences around the date of inspection. In Appendix Table A2 we use 15 days (first three columns) and 30 days (last three columns) after inspection. Columns 1 and 2 show that adding fixed effects for year, month, and day of the week, or calendar date fixed effects, leads to almost identical results. With firm fixed effects (columns 3 and 6) the treatment effects drop from 0.0095 to 0.0069 and from 0.0084 to 0.0053. But firm fixed effects explain part of the variation in voucher use and thus lower the standard errors, leading to very precisely estimated coefficients.

6.3 The 60-minutes SMS Requirement

Next, we analyse the October 2016 SMS reform, which introduced the 60-minute messaging requirement. Prediction 2 in the theoretical model shows that with the SMS requirement (a fall in α), firms with $\theta = \theta^h$ reduce their spot validation of vouchers upon inspection. Consistently to this prediction, firms might decide to reduce or stop misusing vouchers altogether because of the additional and costly constraint of the SMS requirement. Nonetheless, even without on-the-spot validation firms could still be able to exercise the insurance option on a daily level, buying at least one voucher per worker, but this should not anyway generate a jump on the day of the inspection.

Figure 6 shows that once SMS was introduced, there is no more evidence of a substantial and longer-lasting jump.[37] The immediate but very short-term jump (two days) would be consistent with some firms being potentially uninformed about the policy change and trying for one or two days to use vouchers on the spot, before desisting. Overall, only 10 coefficients out of 90 are significantly different from 0 (5 are positive and 5 are negative). Yet, as discussed in Section 3, firms that used vouchers to protect against inspection risk can still do so by buying at least one voucher per worker per day before possible inspections.

Taken together, the evidence supports two distinct conclusions. First, in the pre-SMS regime we observe a sharp inspection-day increase in voucher use, consistent with inspection-contingent misuse (prediction 1 of the model). Second, after the SMS requirement is introduced, this inspection-day discontinuity disappears, indicating that the reform largely reduces or eliminates on-the-spot regularisation through vouchers (prediction 2 of the model).

[37] Given the shorter time period that is available we use -60 to -30 event days as a comparison period.

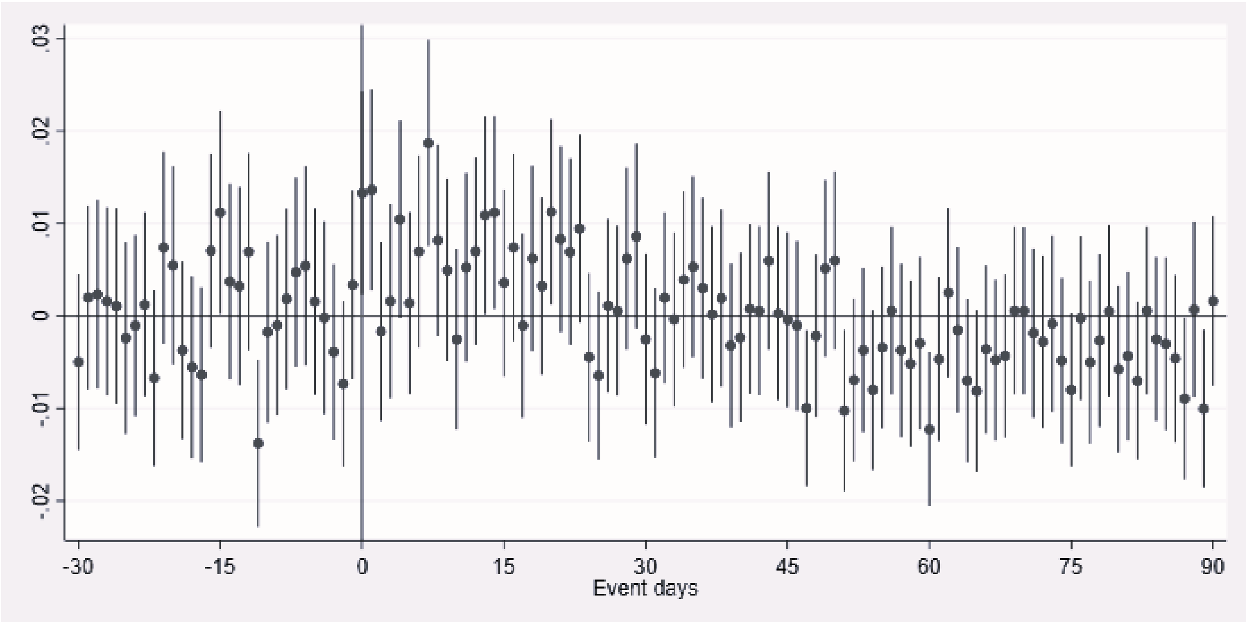


Figure 6: Event Study post-SMS

Notes: The figure plots event study coefficients, where the event is a labour inspection. The excluded time period is between 60 and 30 days prior to the inspection. Standard errors are clustered at the firm level.

Firms may have responded by simply using one voucher per "grey" worker. To test whether the total number of vouchers increased after October 2016, we use monthly data and define as "misusing" firms those who before the SMS requirement and upon inspection were more likely to use vouchers. Figure 7 shows the event-study differences from Equation (12) between "misusing" and "compliant" firms, where 0 refers to the timing of the reform. The monthly estimates are imprecise and should be interpreted cautiously. Point estimates suggest an increase on the order of 10-15 vouchers per firm-month (about 10% relative to baseline), but we cannot rule out zero effects. One plausible reason for the imprecision is heterogeneity in responses, e.g., some firms may have increased ex-ante voucher use while others may have reduced or stopped voucher use altogether. We also detect suggestive pre-trends/anticipation, consistent with the reform having been publicly discussed and potentially known in advance of implementation.[38] In line with the documented slight increase in the number of vouchers, Appendix Figure A7 shows that the SMS requirement led to a slight increase in fixed-term jobs, but without reaching statistical significance.

6.4 The Effect of Vouchers on Regular Employment, Total Wage Bill, and Misclassification Error

6.4.1 The Effect of Vouchers on Regular Employment and Total Wage Bill

We have shown that FWAs that can be activated on the spot can be used to hide undeclared work. In principle, firms could also respond even under information requirements meant to curb this behaviour (such as the SMS rule) by purchasing one voucher per worker per day in advance. The next relevant question is whether vouchers i) crowd out regular work or ii) hide work that would otherwise be fully undeclared (black instead of grey).[39]

[38]Newspaper articles mention this part of the future reform at least since June 2016, see, for example, Fatto Quotidiano (2016).

[39]Given the prevalence in the misuse of vouchers, we are unable to identify whether vouchers give rise to new job opportunities. Although we can use the wage bill as a measure of declared job opportunities and a

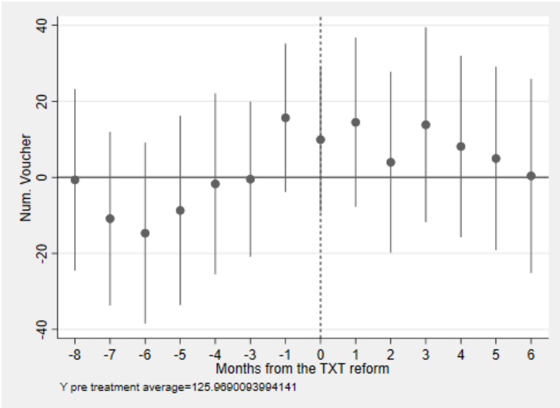


Figure 7: Event Study post-SMS

Notes: The figure plots differences in the number of vouchers used at firms that on average "misused" vouchers and those that did not, 8 months before and 6 months after the abolition of vouchers. Standard errors are clustered at the firm level.

If vouchers were used to cover "grey" jobs, their abolition should induce substitution toward the next most flexible legal contracts, especially among misusing firms. Prediction 3 of our theoretical model predicts that, in the event that FWAs become unavailable, firms should hire some of the "grey" workers using the next most suitable flexible job, and this should occur for low-tax morale firms. The next most FWA is arguably the combination of temporary and part-time labour market contracts.

Using monthly data, we investigate differences in employment outcomes between "misusing" and "compliant" firms in an event-study difference-in-difference approach. Figure 8 shows that when vouchers are abolished, contracts that are at the same time temporary and part-time increase by about 1, representing a 50% increase compared to the pre-abolition average. The difference between "misusing" and "compliant" firms in the total number of temporary and part-time employees is fairly flat in the months leading up to the abolition

lower bound for declared and undeclared job opportunities.

of vouchers in March 2017 and then increases by approximately one additional worker.

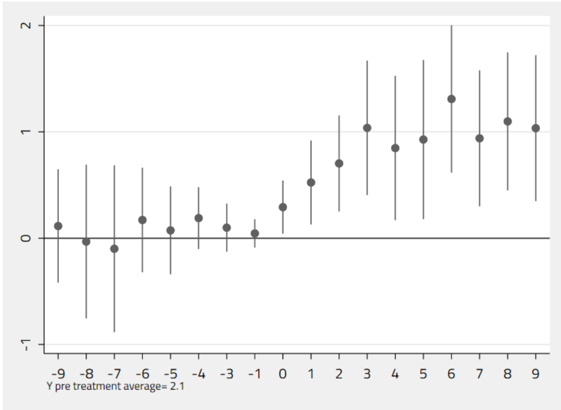


Figure 8: Part-time and Temporary Contracts Around the Abolition of Vouchers

Notes: The figure plots differences in the total number of part-time and temporary workers employed at firms that on average "misused" vouchers and those that did not, 9 months before and 10 months after the abolition of vouchers. Standard errors are clustered at the firm level.

These workers also drive the results for the total number of workers (see Figure 9), since this change is only slightly above 1 (the average number of workers is around 40). Appendix Figure A8 further separates the effects for the two dimensions: temporary vs. permanent (top panel) and part-time vs. full-time (bottom panel). The largest effects are those for temporary workers. For these workers, the difference-in-differences is close to 2, an almost 50% change. There are no effects and, if anything, negative effects for open-ended/permanent contracts (upper right panel). Regarding part-time workers versus full-time workers, both groups show similar effects, indicating that firms seek flexibility with respect to the duration of the labour contracts.

We find no major changes in the declared total wage bill (which includes the cost of vouchers), consistent with substitution across contract types rather than changes in declared

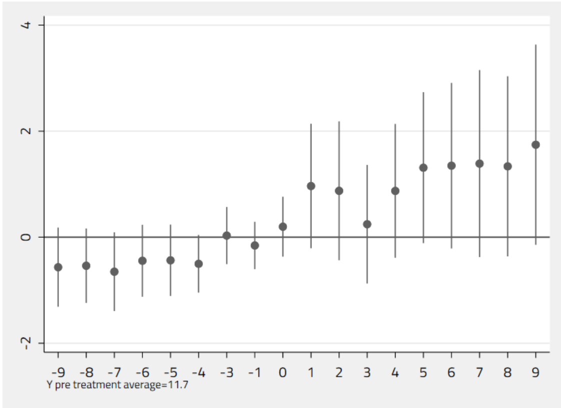


Figure 9: Workers Employed around the Abolition of Vouchers

Notes: The figure plots differences in the total number of workers employed at firms that on average "misused" vouchers and those that did not, 9 months before and 10 months after the abolition of vouchers. Standard errors are clustered at the firm level.

labour costs (see Appendix Figure A9).

Taken together, these results indicate that voucher policy primarily affects the composition of employment (towards temporary/temporary part-time contracts) rather than overall declared labour costs.

6.4.2 Misclassification Error

Because equation (8) identifies only the product $p \times \Delta\tilde{q}^v$, the misclassification exercise is used to tease out the extensive margin p (the share of misusing firms), rather than to identify $\Delta\tilde{q}^v$ on its own.

When defining $\widehat{M}_j^\eta = \mathbf{1}\{\hat{\beta}_j > \eta\}$, with $\eta \geq 0$, the classification of firms as "misusing" depends on the threshold η. A higher η makes the definition more stringent, so it *reduces false positives*—i.e., the probability of classifying a truly compliant firm as misusing (*Type I error*)—but *increases false negatives*—i.e., the probability of failing to classify a truly misusing firm as misusing (*Type II error*). We therefore re-estimate equation (13) while varying η between 0 (our baseline) and 0.075, which mechanically changes the implied share of firms classified as misusing (from roughly 45% at $\eta = 0$ to about 8% at $\eta = 0.075$).

Figure 10 shows that, whether we proxy "true misuse" using temporary contracts only or temporary plus part-time temporary contracts, the threshold η that minimises the sum of Type I and Type II error probabilities is around $\eta \approx 0.03$. This corresponds to classifying approximately 18–19% of firms as misusing in a given day. Relative to the estimates obtained with $\eta = 0$, the corresponding coefficients increase by about 60%, implying that the estimated relative changes in temporary and temporary part-time employment rise from about 50% to more than 75%.

Using our preferred cutoff $\eta \approx 0.03$, about 18% of firms are classified as misusing. Because this classification is subject to Type I/II error, baseline treatment effects should be interpreted as lower bounds. Varying the cutoff η confirms this logic: stricter cutoffs reduce the treated share and increase effect sizes on the intensive margins, consistent with reduced misclassification.

The implied fraction of misusing firms, 18%, is lower than the fraction of firms that are fined at the stage of the final assessment (60%). Importantly, this decomposition refers to misuse at the inspection margin (a point-in-time behaviour), whereas final fines aggregate irregularities detected over the full inspection process.

In terms of the model, it is equivalent to a firm drawing different jobs over time (with different λs) which changes its evading behaviour. As a result, the probability of ever evading over an extended period of time will be larger than the probability of evading at a given moment.

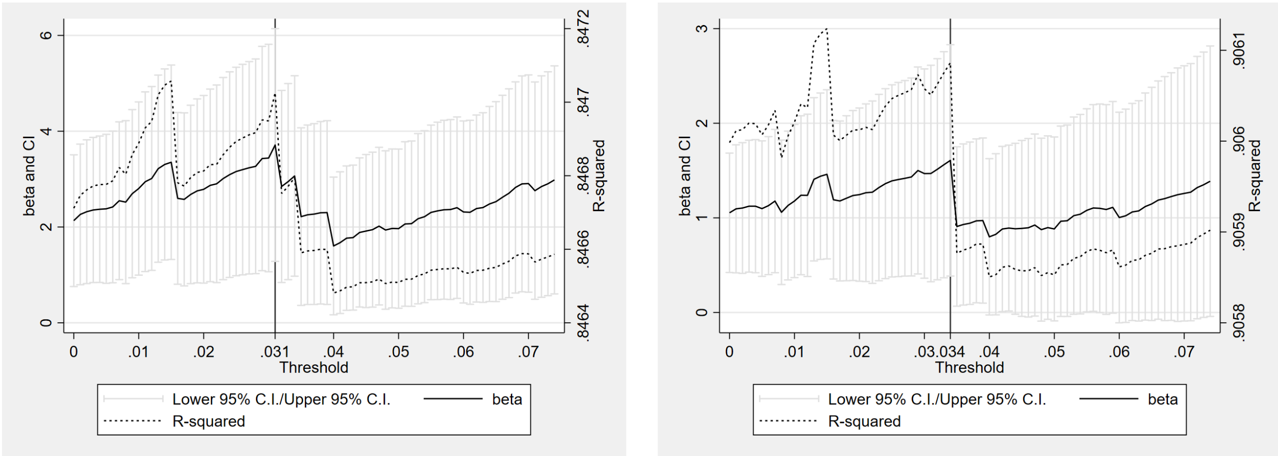


Figure 10: Voucher Abolishment Effects for Different Definitions of Misusing Firms

Notes: The figures plot Difference in Difference estimates for the total number of temporary workers (left panel) and temporary and part-time workers (right panel) between misusing and compliant (and the corresponding 95% confidence intervals) against the η threshold used to separate the two sets of firms: $\widehat{M}_j^\eta = 1\{\hat{\beta}_j > \eta\}$. The dashed line shows the corresponding R-squared. Standard errors are clustered at the firm level.

6.5 The Effect of Vouchers on Evasion

Prediction 3 of our theoretical model predict shadow work to decrease, while due to the absence of the option to hide it with grey contracts, the detected shadow work may actually increase, and this could end up with an increase in fines in case of inspection.

Since the annual probability of inspection is less than 1%, between 2014 and 2017 only a handful of firms were inspected more than once. This implies that we cannot use our usual two-step identification strategy, where we first construct a measure of misuse based on the use of vouchers when inspections are conducted.

We can only compare the amount of social security contributions evaded for inspected firms that have or have not used vouchers between January and October 2016, a clearly less appealing type of comparison.[40] The misuse margin identified in our event-study analysis

[40]We have information on the day the inspection started and the outcome of the inspection (whether a fine was levied and the amount of the related social security contribution evaded).

is immediate, and therefore indicative of evasion at that point in time. By contrast, final assessments reflect the entire investigation period and any additional evidence from firms' records. Separating the inspected firms according to whether they have used at least one voucher is suboptimal, as we have just seen that a minority of firms use vouchers to hide undeclared work. In our sample of inspected firms, 4,269 have used at least one voucher and 16,550 have not, with a total sample size of 20,819 firms.

During the baseline period (January to October 2016), firms that have not used vouchers evade, on average, 26,835 euros. Appendix Table A3 includes the difference-in-difference estimates of a model where time periods are interacted with the treatment variable; that is, firms using vouchers interacted with time periods (Pre-SMS; from 1 May to 16 October 2016, the omitted category; Post-SMS, from 17 October 2016 to 17 March 2017; Post-Abolition, from 18 March to 31 December 2017). Four different specifications are reported, where column (1) does not include any additional control variables to the baseline difference-in-difference, while columns (2) to (4) include several sets of controls.

It emerges that, compared to control firms, firms that used vouchers tend to evade more (i.e. pay higher fines when inspected) when vouchers become unavailable. The results change little when we control for a monthly cubic time trend, and for employment characteristics (the share of part-time workers, fixed-term workers, their interaction, and wages). For the SMS requirement period, the effect is negative, though non-statistically different from zero. The model implies offsetting channels: the SMS requirement may reduce misuse/evasion, but at the same time it might raise detection, and the net effect is ambiguous.

7 Conclusions

In response to the increasing demand from firms for flexible work contracts, legislators around the world are devising labour contracts that allow firms to hire workers for just a small amount of work, i.e., on-call work, zero-hour work, labour vouchers, mini-jobs, etc.

These contracts, while offering flexibility, often result in poor career development prospects, stagnant wages, and increased exposure to income risk, as workers find themselves in precarious employment situations without stable income or career growth opportunities (Boeri et al., 2020). This study documents an additional important risk: these labour contracts may complicate the job of labour inspectors whose task is to uncover undeclared work. We show that upon random inspections, whenever it is lawful to use labour vouchers on the spot, some firms use FWAs to hide undeclared work. Both the labour demand model and the evidence show that this novel mechanism can counteract the expected reduction in undeclared work resulting from lower hiring and firing costs.

Our model and evidence also show that enforcement effectiveness depends on institutional design. Stronger enforcement deters shadow work, but this deterrence is attenuated when firms can regularise on the spot through vouchers. In this sense, policies that make instant relabelling easier can weaken the impact of inspections, even when formal enforcement intensity is unchanged.

Therefore, alongside limits on voucher amounts, policymakers should prioritise ex-ante traceability rules and narrow eligibility criteria, as these design features preserve flexibility while restoring deterrence.

We estimate that 18% of firms used FWAs to hide undeclared work, and that these firms hired more fixed-term and part-time workers when FWA were banned. In addition, the ban increased average inspection fines for firms that used vouchers before they were banned.

Our evidence from Italy suggests that when highly flexible work arrangements can be validated on the spot, they may serve as an inspection-contingent regularisation device, complicating enforcement and potentially increasing undeclared work in similar institutional settings. In addition, firms that use vouchers to hide undeclared work tend to use vouchers mainly when inspected. As a result, restricting the total number of vouchers used by workers or firms to be below some threshold, which has been Italy's primary policy restriction, is unlikely to successfully curb misuse. Rules that restrict the use of FWA to certain categories, such as students or retirees, are probably a better way to balance the need for flexibility with legality and the protection of employment rights.

References

Abi Adams and Jeremias Prassl. Zero-hours work in the United Kingdom. Technical Report Conditions of Work and Employment Series No. 101, International Labour Organization, 2018.

James Albrecht, Lucas Navarro, and Susan Vroman. The Effects of Labour Market Policies in an Economy with an Informal Sector. *The Economic Journal*, 119(539):1105–1129, 06 2009a. ISSN 0013-0133.

James Albrecht, Lucas Navarro, and Susan Vroman. The effects of labour market policies in an economy with an informal sector. *The Economic Journal*, 119(539):1105–1129, 2009b.

Susan Athey and Guido W. Imbens. Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79, 2022. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2020.10.012. URL `https://www.sciencedirect.com/science/article/pii/S0304407621000488`. Annals Issue in Honor of Gary Chamberlain.

Aniko Biro, Daniel Prinz, and Laszlo Sandor. The minimum wage, informal pay, and tax enforcement. *Journal of Public Economics*, 215:104728, 2022. doi: 10.1016/j.jpubeco.2022.104728.

Olivier Blanchard and Pedro Portugal. What hides behind an unemployment rate: Comparing portuguese and us labor markets. *American Economic Review*, 91(1):187–207, 2001.

Tito Boeri and Pietro Garibaldi. Temporary Employment in Markets with Frictions. *Journal of Economic Literature*, 62(3):1143–1185, 2024.

Tito Boeri, Giulia Giupponi, Alan B Krueger, and Stephen Machin. Solo self-employment and alternative work arrangements: a cross-country perspective on the changing composition of jobs. *Journal of Economic Perspectives*, 34(1):170–95, 2020.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, 91(6):3253–3285, 02 2024. ISSN 0034-6527. doi: 10.1093/restud/rdae007. URL `https://doi.org/10.1093/restud/rdae007`.

Mariano Bosch and Julen Esteban-Pretel. Job creation and job destruction in the presence of informal markets. *Journal of Development Economics*, 98(2):270–286, 2012.

Pierre Cahuc, Olivier Carlot, and Frank Malherbet. Explaining the spread of temporary contracts and its impact on labor turnover. *The International Economic Review*, 57(2): 533–572, 2016.

David C. Chan. The efficiency of slacking off: Evidence from the emergency department. *Econometrica*, 86(3):997–1030, 2018.

M Keith Chen, Peter E Rossi, Judith A Chevalier, and Emily Oehlsen. The value of flexible work: Evidence from UBER drivers. *Journal of Political Economy*, 127(6):2735–2794, 2019.

Nikhil Datta, Giulia Giupponi, and Stephen Machin. Zero-hours contracts and labour market policy. *Economic Policy*, 34(99):369–427, 2019.

Alessandro Del Vecchio. La ragionevole durata dell'ispezione in materia di lavoro e previdenza. *Rivista del Diritto della Sicurezza Sociale*, 19(4):813–832, 2019.

Edoardo DiPorto, Leandro Elia, and Cristina Tealdi. Informal work in a flexible labour market. *Oxford Economic Papers*, 69(1):143–164, 2017.

Il Fatto Quotidiano. Voucher lavoro, renzi blocca decreto con i correttivi troppo soft. “timore di attacchi da opposizioni e sindacati”. *Il Fatto Quotidiano*, June 1 2016.

Marco Fugazza and Jean-François Jacques. Labor market institutions, taxation and the underground economy. *Journal of Public Economics*, 88(1-2):395–418, 2004.

Nicolas Gavoille and Anna Zasova. What we pay in the shadows: Labor tax evasion, minimum wage hike and employment. *Journal of Public Economics*, 228:105027, 2023. doi: 10.1016/j.jpubeco.2023.105027.

Andrew Goodman-Bacon. Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 2021.

ILO. Labour Inspection and Undeclared Work in the EU . Technical report, ILO, 2013.

Simon Johnson, Daniel Kaufmann, and Pablo Zoido-Lobaton. Corruption, Public Finances and the Unofficial Economy. *Policy Research Working Paper*, (2169), 1998.

Lawrence F Katz and Alan B Krueger. The rise and nature of alternative work arrangements in the united states, 1995–2015. *ILR Review*, 72(2):382–416, 2019.

William F Maloney. Informality revisited. *World development*, 32(7):1159–1178, 2004.

Irene Mandl. New forms of employment: 2020 update. Technical report, The European Foundation for the Improvement of Living and Working Conditions, Publications Office of the European Union, Luxembourg, 2020.

Alexandre Mas and Amanda Pallais. Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–59, 2017.

Jonathan A Parker, Nicholas S Souleles, David S Johnson, and Robert McClelland. Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, 103(6):2530–53, 2013.

Friedrich Schneider and Dominik H Enste. Shadow economies: Size, causes, and consequences. *Journal of economic literature*, 38(1):77–114, 2000.

Mr Friedrich Schneider and Dominik Enste. *Hiding in the shadows: the growth of the underground economy.* International Monetary Fund, 2002.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 2020.

Mirco Tonin. Minimum wage and tax evasion: Theory and evidence. *Journal of Public Economics*, 95(11-12):1635–1651, 2011. doi: 10.1016/j.jpubeco.2011.04.005.

Gabriel Ulyssea. Firms, informality, and development: Theory and evidence from brazil. *American Economic Review*, 108(8):2015–47, 2018.

Gabriel Ulyssea. Informality: Causes and consequences for development. *Annual Review of Economics*, 12:525–546, 2020.

Colin Williams, Predrag Bejaković, Davor Mikulić, Josip Franic, Abbi Kedir, and Ioana Alexandra Horodnic. An evaluation of the scale of undeclared work in the european union and its structural determinants: estimates using the labour input method. *European Commission, Directorate-General for Employment, Social Affairs and Inclusion, Publications Office of the European Union, Luxembourg*, 2017.

Colin W. Williams. Elements of a preventative approach towards undeclared work: an evaluation of service vouchers and awareness raising campaigns. Technical report, European Platform Undeclared Work, 2018.

A Appendix

A.1 The Reservation Values

The threshold λs are:

$$\begin{cases} \lambda^{ft} & = \frac{\rho^{ft}\tilde{y}}{\rho^{ft}\tilde{y}+f\omega-(1-\rho^{ft})(\omega+\tau)} \\ \\ \lambda^{v} & = \frac{(\rho^{v}-\rho^{ft})\tilde{y}}{(\rho^{v}-\rho^{ft})\tilde{y}+(1-\rho^{ft})(\omega+\tau))} \end{cases} \tag{16}$$

Note that existence of two thresholds-and thus two different contracts with fixed duration (vouchers and fixed term contracts)-requires that the duration of FWAs is sufficiently short or that

$$\rho^{v} > \frac{\rho^{ft}F}{F+(1-\rho^{ft})(w+\tau)}.$$

A.2 Proof of Proposition 2

The decision to go shadow for FWA jobs requires that for $\gamma c > (1-\gamma)(1-\rho^{FWA})\tau + \theta^{h}$ which implies $\theta^{h} < \gamma c - ((1-\gamma)(1-\rho^{FWA})\tau)$ or $c > \frac{((1-\gamma)(1-\rho^{FWA})\tau)}{\gamma}$. Further, $\theta^{h} < \gamma c - (1-\rho^{FWA})(1-\gamma)\tau > 0$. With these restrictions, the marginal shadow value is

$$\tilde{\lambda}_{s,\mathrm{FWA}}(\theta_h) = 1 - \frac{\theta_h}{D}, \qquad D \equiv \gamma\, c\,(1-\alpha) - (1-\gamma)\,\tau\,(1-\rho^{FWA}) \;>\; 0, \tag{17}$$

which is lower than λ^{s} as long as θ^{h} satisfies the condition above.

A.3 How grey regularisation weakens the role of enforcement.

From the misuse condition in (6), define

$$S(\lambda;\gamma,\alpha) \equiv (1-\gamma)\big(J^{h}(\lambda)-J^{*}(\lambda)\big)+\theta-\gamma\big[(1-\alpha)C(\lambda)+\alpha\big(J^{*}(\lambda)-J^{v}(\lambda)\big)\big]\,.$$

A job is operated in shadow/misuse mode when $S(\lambda;\gamma,\alpha) > 0$. The direct effect of enforcement is

$$\frac{\partial S}{\partial \gamma} = -\big(J^h(\lambda) - J^*(\lambda)\big) - \big[(1-\alpha)C(\lambda) + \alpha\big(J^*(\lambda) - J^v(\lambda)\big)\big] < 0,$$

so stronger enforcement ($\gamma \uparrow$) reduces the incentive to remain shadow. To see how this effect varies with the grey-regularisation technology α, compute

$$\frac{\partial^2 S}{\partial \alpha\, \partial \gamma} = C(\lambda) - \big(J^*(\lambda) - J^v(\lambda)\big).$$

Hence, under the empirically plausible condition

$$C(\lambda) > \big(J^*(\lambda) - J^v(\lambda)\big),$$

we have $\frac{\partial^2 S}{\partial \alpha\, \partial \gamma} > 0$: as α increases (on-the-spot regularisation becomes easier), the deterrent effect of γ becomes weaker. Intuitively, when firms can more easily convert detected shadow work into vouchers at inspection, a higher inspection probability carries less effective punishment, so enforcement shifts total shadow work by less.

A.4 Tables and Figures


PERIODO PRESTAZIONE
dal ..... /..... /..... al ..... /..... /.....
CODICE FISCALE DATORE DI LAVORO
CODICE FISCALE LAVORATORE
FIRMA LAVORATORE
>700001003102472<
>700001003102472<
INPS
INAIL
BUONO LAVORO
Prestazione di lavoro occasionale di tipo accessorio
Valido per prestazioni effettuate entro il 31/12/2008
Valore netto EURO 7,50

Figure A1: The Photograph of an Italian voucher

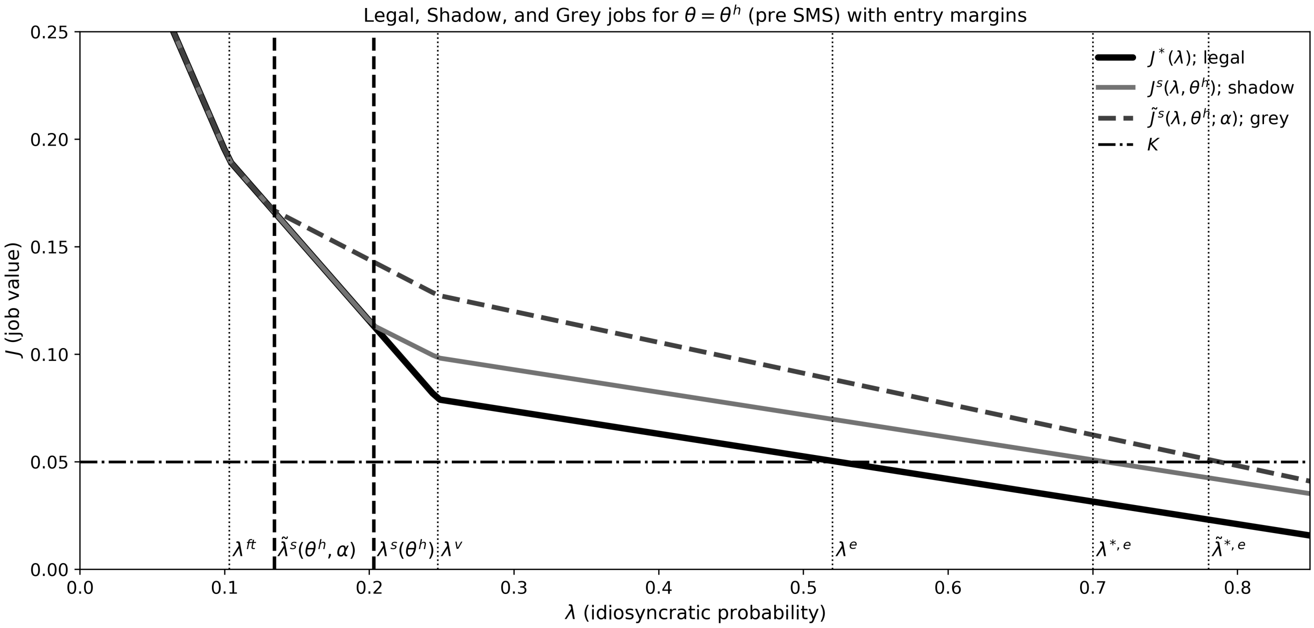


Figure A2: Optimal Contracts and Shadow Employment For Firms with High Tax-evasion Attitude, including Entry Margins.

Notes: The chart reports the value functions for legal jobs $J^*(\lambda)$ and shadow jobs with misuse $\tilde{J}^s(\lambda, \theta^j)$. Parameters are $y = 1, \omega = .4, \tau = .25, f = 3$, and $\rho^{ft} = .23, \rho^v = .7, K = 0.05$. The chart considers also the shadow option with linear cost sanction $C = c(1 - \lambda)$ with $c = .92$. $\theta^h = 0.019$ and $\theta^l = -0.081$. The thresholds λ^{ft}, λ^v, λ^s, $\tilde{\lambda}^s$ are indicated in the chart. In the top panel the bold continuous piecewise linear function refers to legal jobs ($J^*(\lambda)$), while the lighter continuous function refers to black shadow jobs for firms with $\theta = \theta^h$. Firms with $\theta = \theta^h$ go shadow for $\lambda > \lambda^s(\theta^h)$. The margins λ^e, $\lambda^{*,e}$, and $\tilde{\lambda}^{*,e}$ refer to the entry margins for Legal jobs, Black Shadow Jobs and Grey Shadow Jobs.

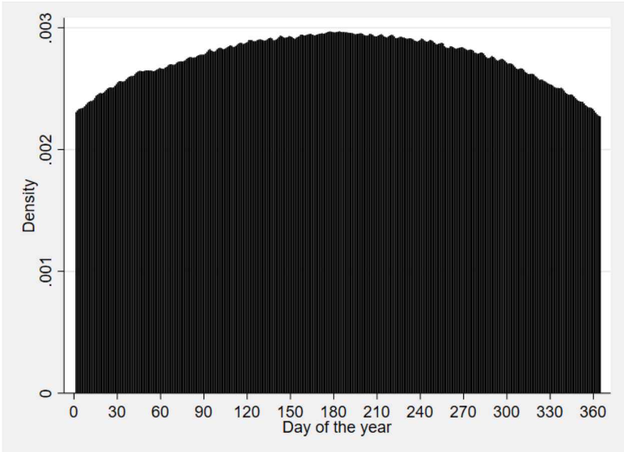


Figure A3: Distribution of Labour Inspections

Notes: The figure plots the histogram of the exact day of inspection.

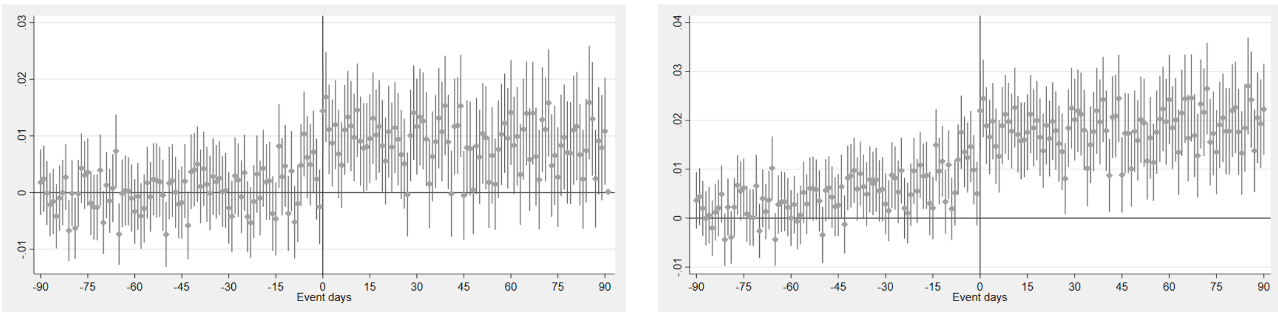


Figure A4: Event Study pre-SMS

Notes: The figure plots event study coefficients, where the event is a labour inspection. The excluded time period is between 180 and 90 days prior to the inspection. Standard errors are clustered at the firm level. The figure on the left controls for calendar time t (in Eq. 9, $f(t) = \alpha t$), the figure on the right has no additional controls (in Eq. 9, $f(t) = 0$).

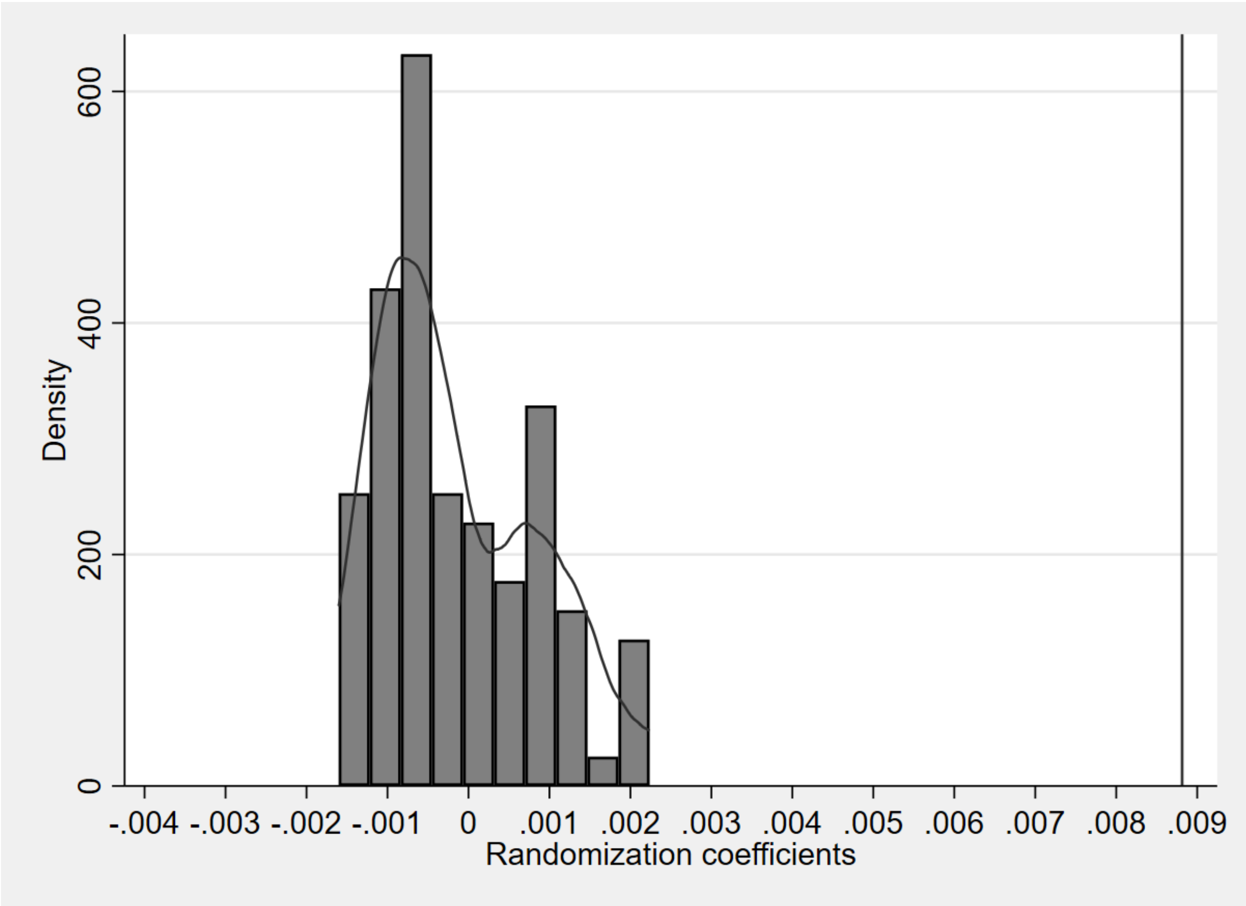


Figure A5: Randomization Test

Notes: The figure plots the density of the randomization coefficients in a simple difference model: $DV_{j,\tau} = \alpha + \beta D_{\tau-t} + \epsilon_{j,\tau}$ where $t < 0$ measures the event time of inspection (it's zero the day of the inspection), while $D_{\tau-t}$ takes value one t days before the inspection. The vertical line on the right corresponds to the coefficient when the timing of the labour inspection is correct ($t = 0$) and the full sample is used ($t \lesseqgtr 0$).

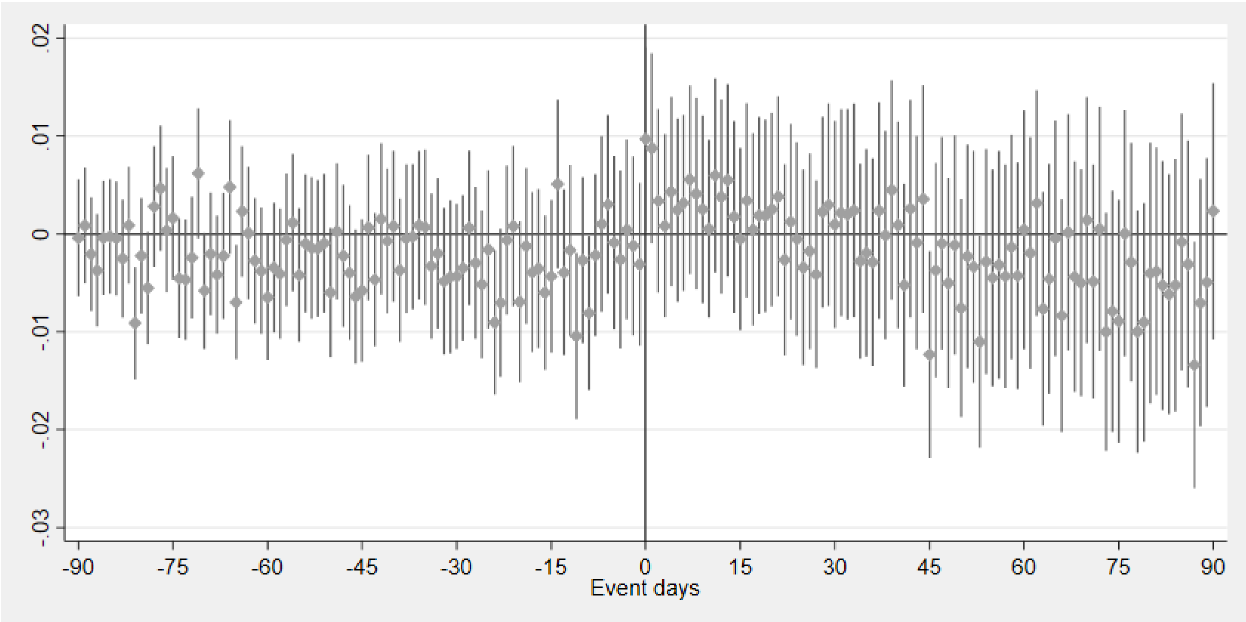


Figure A6: Event Study pre-SMS with Firm Fixed Effects

Notes: The figure plots event study coefficients, where the event is a labour inspection. The excluded time period is between 180 and 90 days prior to the inspection. The regression controls for date and firm fixed effects. Standard errors are clustered at the firm level.

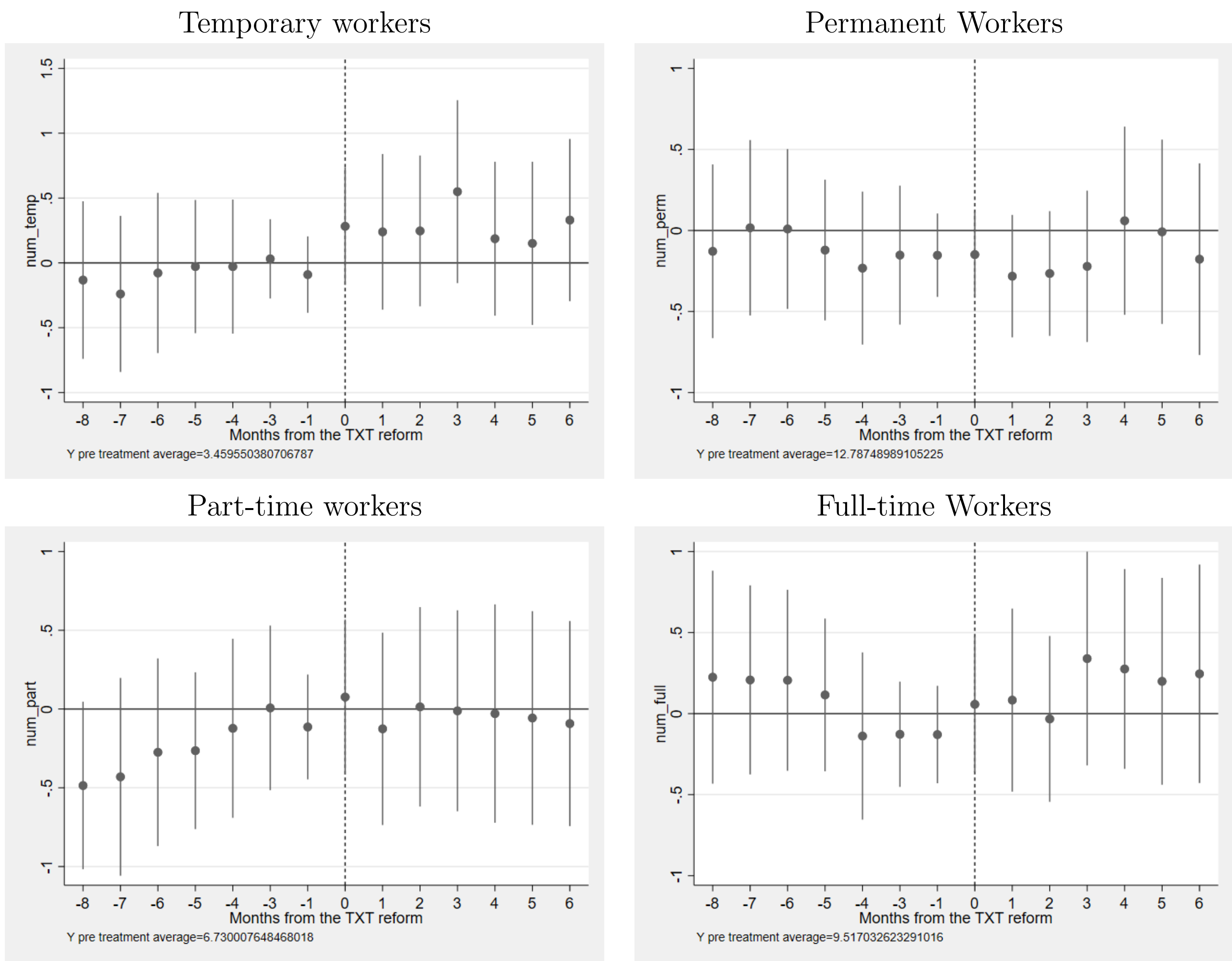


Figure A7: Event Study Around the Introduction of the SMS Requirement

Notes: The figure plots differences in the number of workers employed at firms that on average "misused" vouchers and those that did not, 9 months before and 6 months after the Introduction of the SMS (we have to limit the analysis to six months after as in March 2017 vouchers were abolished). Standard errors are clustered at the firm level.

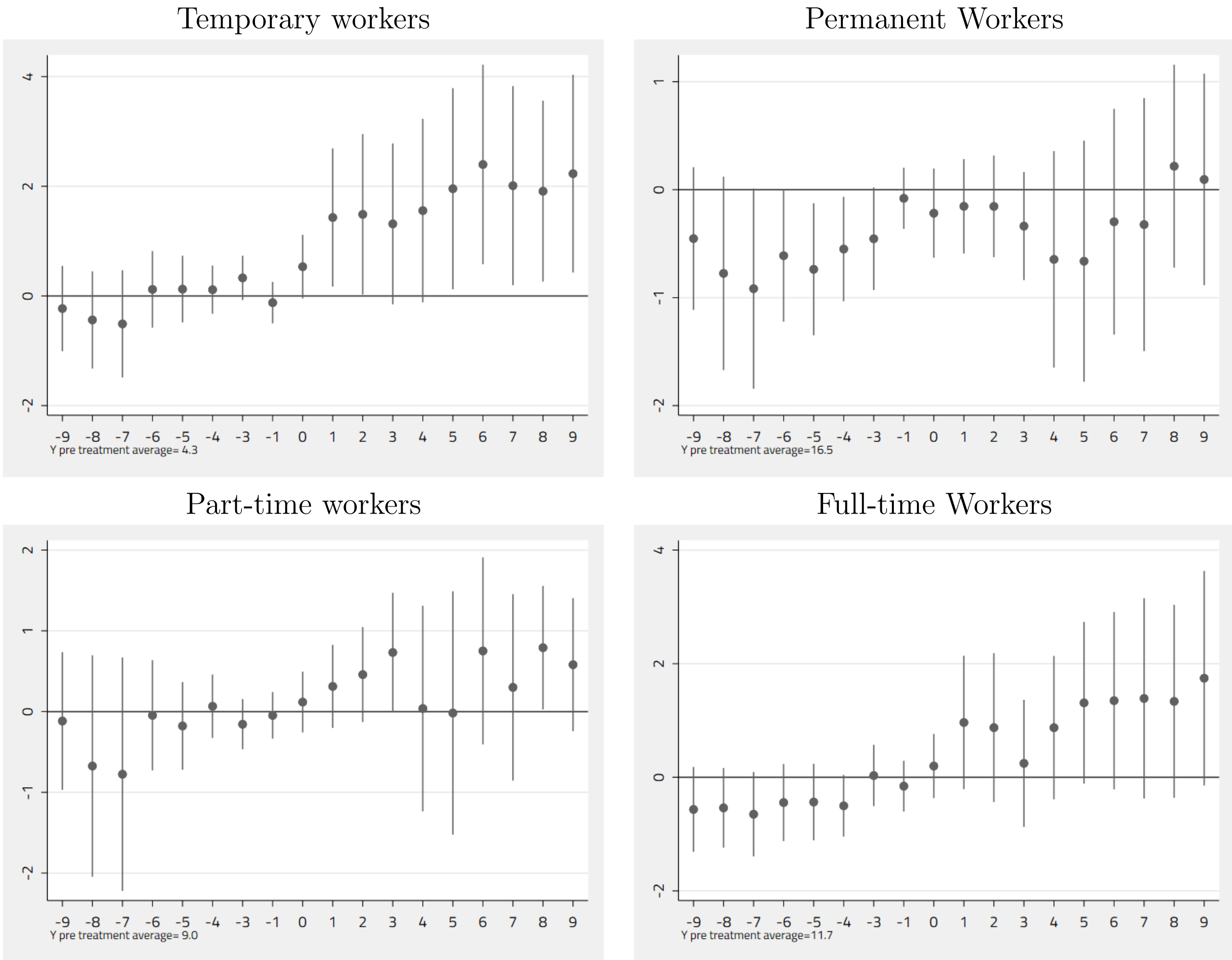


Figure A8: Event Study around the Abolition of Vouchers

Notes: The figure plots differences in the number of workers employed at firms that on average "misused" vouchers and those that did not, 9 months before and 10 months after the abolition of vouchers. Standard errors are clustered at the firm level.

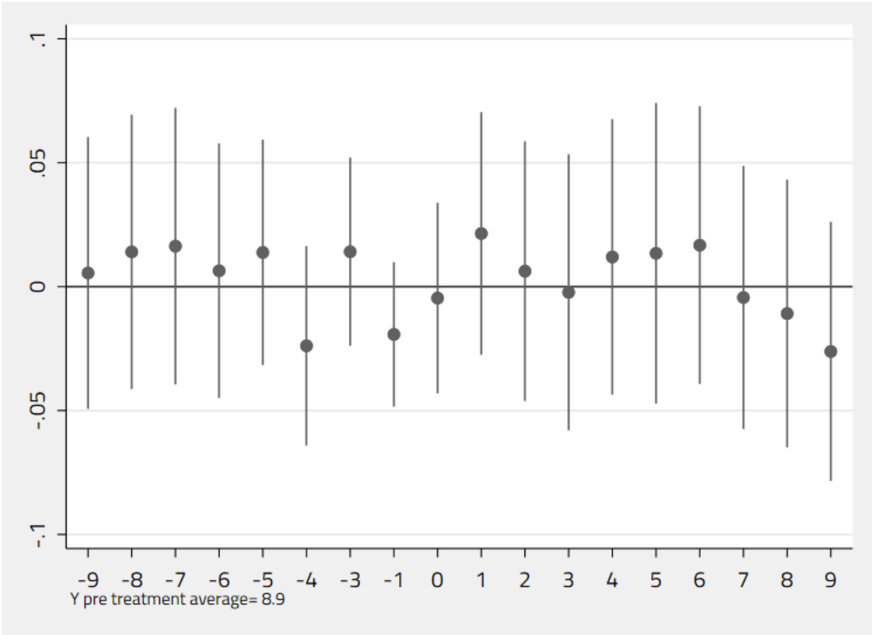


Figure A9: Log-Total Wage Bill Around the Abolition of Vouchers

Notes: The figure plots differences in the total wage bill between firms that on average "misused" vouchers and those that did not, 10 months before and 9 months after the abolition of vouchers. Standard errors are clustered at the firm level.

Table A1: Vouchers and labour Inspections Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Subsample	South	Center	North-East	North-West	Young Firms	Medium-aged F.	Old Firms	Small Firms	Medium-sized F.	Large Firms	High PT Use	Above 20% Temporary	Below 20% Temporary
Post-Inspection	0.009**	0.008**	0.010***	0.011***	0.009***	0.011***	0.008**	0.006***	0.008***	0.012***	0.014***	0.011***	0.085***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
Constant	0.047***	0.046***	0.047***	0.043***	0.045***	0.048***	0.044***	0.035***	0.044***	0.058***	0.057***	0.052***	0.039***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	255,262	256,735	409,649	347,459	581,120	414,932	273,053	461,643	373,953	433,509	446,903	644,001	534,543
R-squared	0.015	0.015	0.014	0.014	0.014	0.013	0.017	0.012	0.016	0.017	0.005	0.015	0.013
Mean dep. var.	0.0495	0.0477	0.0504	0.0461	0.0477	0.0498	0.0481	0.0368	0.0466	0.0609	0.0574	0.0548	0.0421

Notes: Linear probability model of using at least one voucher, daily data with year, month and day of the week fixed effects. Clustered standard errors (by firm) in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Small firms have less than five employees, Medium-sized firms between 5 and 15, while Large firms are those with more than 15 employees: old firms are those with more than 14 years of age, medium are between 5 and 14 years of age, and young firms are below 5 years of age. High PT use refers to an incidence of Part Time above the median.

Table A2: Difference-in-Differences Models of Vouchers and Labour Inspections

	(1)	(2)	(3)	(4)	(5)	(6)
Post period		15 days			30 days	
Post-Inspection	0.0094***	0.0095***	0.0069***	0.0083***	0.0084***	0.0053***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Firm Fixed Effects	No	No	Yes	No	No	Yes
Time Fixed Effects	No	Yes	Yes	No	Yes	Yes
Year, month, dow Fes	Yes	No	No	Yes	No	No
Observations	713,849	713,779	713,773	761,100	761,015	761,012
R-squared	0.023	0.025	0.165	0.023	0.025	0.164

Notes: Linear probability model of using at least one voucher with daily data. Clustered standard errors (by firm) in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Vouchers and Evaded Contributions

	(1)	(2)	(3)	(4)
Post-Abolition*Treated	30,562**	32,615**	34,465**	32,876**
	(14,835)	(14,654)	(16,503)	(16,566)
Post-Sms*Treated	-7,076	-6,732	-6,272	-7,628
	(6,726)	(6,714)	(7,564)	(7,559)
Treated	-3,640	-3,737	-4,386	-4,032
	(3,635)	(3,621)	(4,058)	(4,070)
Post-Abolition	12,426**	10,202*	11,474*	10,756*
	(5,806)	(5,573)	(6,085)	(6,001)
Post-Sms	7,334	3,021	737	1,134
	(4,834)	(5,655)	(6,361)	(5,219)
Observations	20,817	20,817	18,108	18,108
Controls for Employment	No	No	No	Yes
Controls for Wages	No	No	Yes	Yes
Cubic Monthly Trend	No	Yes	Yes	Yes

Notes: The dependent variable is the amount of evaded contributions for each inspected firm. Treated firms have used at least one voucher during the sample period, control firms have never used vouchers. The baseline period is from January to October 17, 2016, while the Post-SMS period goes from October 17, 2016 to March 17, 2017 (SMS introduction), and the Post-Abolition period starts in March 17, 2017 (Abolition date) to the end of 2017. Standard errors are clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.