The Perverse Effect of Flexible Work Arrangements on Informality*

Edoardo Di Porto[†] Pietro Garibaldi[‡] Giovanni Mastrobuoni[§¶]

Paolo Naticchioni[‖]

October 23, 2024

Abstract

Flexible work arrangements are increasingly common in many countries, as they allow for quick adjustments in labor demand. These arrangements are also thought to discourage undeclared work, although the evidence is mostly correlational. Using Italian administrative data on labor vouchers and randomly timed labor inspections, this study demonstrates that flexible work arrangements actually disrupt the work of labor inspectors, leading to more undeclared work rather than less. Firms using flexible work arrangements to hide undeclared work tend to hire more regular part-time and fixed-term workers and are more likely to be fined by labor inspectors when vouchers are abolished. A simple partial equilibrium labor demand model rationalizes these findings.

Keywords: informality, labor vouchers, flexible work arrangements, occasional work, zero-hour contracts.
JEL codes: J23, H26.

*We thank Bruno Anastasia, Saverio Bombelli, Lorenzo Cappellari, Tullio Jappelli, Marteen Lindeboom, Roberto Nisticó, 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. 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.

[†]University of Naples "Federico II" and CSEF - UCFS Uppsala University, edoardo.diporto@unina.it

[‡]Collegio Carlo Alberto, University of Torino, CEPR, IZA, pietro.garibaldi@unito.it

[§]Collegio Carlo Alberto, University of Torino, CEPR, IZA, giovanni.mastrobuoni@carloalberto.org

[¶]Corresponding author.

[‖]Roma Tre University - IZA, paolo.naticchioni@uniroma3.it

1 Introduction

In most advanced countries, it is estimated that 10 to 20% of the economy is underground (Schneider and Enste, 2002).[1] Furthermore, undeclared work is very persistent and, if anything, its share has been estimated to increase with time (Ulyssea, 2018).

In response, as also mentioned by the ILO (2013),[2] governments have introduced more flexible labor contracts, known as alternative work arrangements (AWAs).[3] Approximately 90% of European countries have arrangements with no guaranteed working hours, allowing firms to quickly adjust labor demand and providing workers with more flexible schedules.[4]

Although there is little evidence on the relationship between AWAs and undeclared work, many experts believe that this relationship should be negative.[5] For example, the European Union Agency for the Improvement of Living and Working Conditions mentions that some new forms of employment have been developed to help formalize undeclared work practices.[6] The European Platform Tackling Undeclared Work has also stated that both Social and Enterprise Voucher schemes should target areas where undeclared work is prevalent (see Williams, 2018).

This paper aims to understand whether flexible work schedules influence firms' decisions to hire undeclared workers. We posit that flexible work schedules may interfere with labor inspectors' ability to detect undeclared work. The types of contracts we have in mind are,

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

[2]According to the ILO, Belgium, France, Finland, Denmark, and Switzerland, introduced AWAs to discourage undeclared work.

[3]For an overview on this early stage literature see Boeri et al. (2020), while Katz and Krueger (2019) cover AWAs for the US and Adams and Prassl (2018) and Datta et al. (2019) for Europe.

[4]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.

[5]There is also evidence that rigid employment protection legislation reduces job flows and pushes 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 labor 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.

for example, UK's "Zero Contract Hours," where workers may work more than the officially declared number of hours or Italy's labor vouchers, where for a day of work a worker may receive a single voucher for his work, so as to justify his or her physical presence in the workplace, and be paid the rest under the table. According to a report by the Ministry of Labor, issued when it reformed some AWA rules in 2016, some firms may behave like "a citizen who validates his bus ticket only when the ticket inspector gets on the bus"[7] Interestingly, the report also provides evidence that about 10% of AWA workers were previously employed by the same firm but with more regular contracts.

We model firms' labor demand for different contract types, allowing for this mechanism. When this mechanism is turned off, more flexible job contracts reduce hiring and/or firing costs and lower the demand for undeclared work.[8] When we turn it on, allowing contracts to have no work schedule, firms simply report the working hours worked by their casual employee only when they are inspected.

Labor force surveys often lack the detail to identify AWA arrangements (Katz and Krueger, 2019), and when paid sums do not contribute to future social security benefits, even administrative data can be uninformative. Additionally, as shown by Mas and Pallais (2017), workers often select into AWAs, complicating the estimation of counterfactual scenarios.

We exploit data from Italy, a country with a sufficiently large underground economy and a legislation that has first liberalized and later abolished AWAs. Within the European Union, only Eastern European countries, Spain and Greece, have a higher prevalence of undeclared work[9] and beginning about 15 years ago, employers could purchase 10-euro vouchers to pay for work without needing a formal labor contract. The worker would later exchange vouchers for money. This system was intended to discourage undeclared work by eliminating

[7]See Ministero del Lavoro (2016).
[8]See Albrecht et al. (2009a), Bosch and Esteban-Pretel (2012), and Ulyssea (2018).
[9]According to Williams et al. (2017), 17.2% of Italian work in undeclared (the EU average is 16.4%).

bureaucracy and reducing hiring and firing costs.[10]

Descriptive evidence, included in a report produced by the Italian Social Security Administration (Anastasia et al., 2016), appeared to support this view, although the evidence was correlational and may have been driven by differences in economic activity: economic growth is likely to increase the demand for vouchers and reduce the willingness of workers to work under the table.[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 purchases of vouchers between 2014 and 2017; and finally, iii) data on the universe of labor inspections between 2014 and 2017. Our data reveal that almost one firm out of four in 2016 was using vouchers. Using the unpredictability of the timing of labor inspections, which we document, we find clear evidence that–as soon an inspection starts–some firms tend to immediately increase their use of vouchers.[12] 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%).

Next, we test whether AWAs displace regular work. Dividing inspected firms into those that upon inspection, on average, increased their use of vouchers and those that did not,

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

[11]In 2015, regions where the average number of vouchers per worker was higher tended to have fewer undeclared work. The two extreme cases were the regions of Lombardy and Calabria. The rich northern region had an average of 78 vouchers per casual worker and was estimated to have less than 10% undeclared work. The poor southern region, where one in four workers was estimated to be undeclared, used about half as many vouchers per casual worker. Although the implied elasticity is close to -60%. The elasticity using the South as the baseline is $(0.25 - 0.10)/0.39 \times 0.39/0.25$.

[12]Moreover, such a jump ceases to exist when the government introduces a small change in legislation, requiring firms to announce the use of vouchers with at least one hour notice. The one hour notice gave inspectors enough time to uncover undeclared work, forcing firms to stop using vouchers on the spot. After this policy change, firms that were previously misusing AWAs start to buy more vouchers, regardless of inspections. It appears that simply buying a voucher per worker per day ensures that firms remain insured against the risk of inspections.

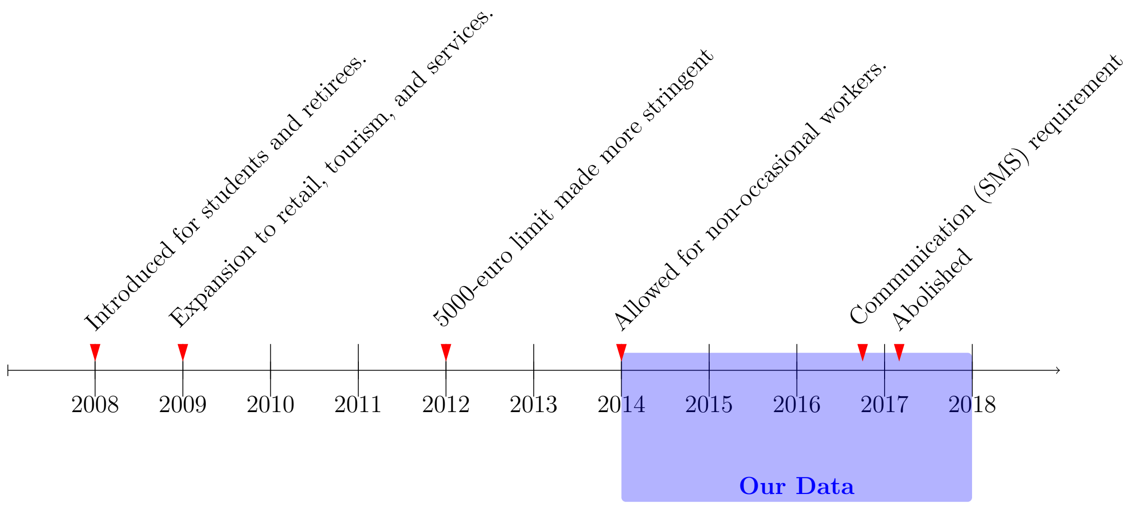
we analyze what these two sets of firms did in March 2017, when vouchers were indeed abolished. Although our measure of misbehaving firms is subject to misclassification error, as changes in individual-level use of AWAs around labor inspections may reflect unobservable demand shocks, we show that our main estimates are lower bounds of the true effects and that varying the definition of misbehavior, and thus the size of the effects can be informative about the fraction of firms hiding undeclared work.

We show that about 18% of the inspected firms that use vouchers misuse AWAs. In line with the predictions of the model, "misusing" firms are shown to revert to the next most flexible work contracts, hiring approximately two additional fixed-term workers, representing a 50% increase with respect to the preabolishment average of misusing firms. Due to these substitution effects, the total declared wage bill, which includes the cost of vouchers, shows no changes, a possible rule-of-thumb behavior of marginal firms.

To evaluate the impact of voucher abolition on undeclared work, we compare the amount of evaded social security contributions for firms that have (treated) or have not (control) used any vouchers between January and October 2016: treated firms tend to evade more when vouchers were abolished, while at the height of the use of vouchers (right before and after the introduction of the SMS requirement) evasion is on average lower.

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

Figure 1: Timeline of Voucher Legislation



2 The Institutional framework

Figure 1 outlines the timeline from when AWAs were first introduced to when they were abolished. In 2008 the Italian legislator introduced AWAs in the extreme form of labor vouchers. Employers could buy vouchers from the Social Security Administration (INPS), or in tobacco stores, banks, and post offices. The vouchers looked like checks (see Appendix Figure A5), and employers who want to pay someone for completing a temporary job would fill them with the worker's social security number "*Codice Fiscale*" and date and hand them over to the worker. The workers would be able to easily cash these vouchers. For every 10 euros paid by the employer, the worker received 7.50 euros, 1.30 euros covered the social security contributions (less than the typical labor contract, where contribution rates are 1/3), 70 cents the health insurance and 50 cents the commission fee paid to the social security administration. Firms were allowed to buy vouchers in bulk, and they did, but the data contain information on the date vouchers were actually used.

Initially, vouchers had considerable restrictions: employers could only spend a maximum of 5000 euros in vouchers for each employee; only students and retirees were allowed to receive vouchers and only in the agricultural sector.

Several small changes in the initial conditions led to a steep increase in the use of AWAs. Initially, the center-right government extended 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 3, this led to rapid growth in the monthly number of 10-euro vouchers sold: from a few thousands in 2008 to a peak of almost 20 million in 2016.

In 2009 vouchers became available in the retail sector, tourism, and service sector, and for domestic workers. One year later, they were completely liberalized, opening up to all sectors and all workers. After a temporary setback in 2012, when the worker's 5000-euro limit was made more stringent, as it applied to the sum across all the employers and not to each employer separately, the 2014 labor reform allowed vouchers not to be related to occasional work, and their annual limit increased to 7000 euros.

The use of vouchers reached a peak in 2016, when the pressure from labor unions to reform their use or completely abolish them intensified. In October 2016, a first law reform was implemented, as companies 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 took over, the vouchers were completely removed.[13]

Inspections occur at random and unpredictable times and their duration can last from a few days to several weeks. Inspectors are supposed to uncover nondeclared work and hazardous working conditions. Depending on the outcome of the inspection, companies may be required to pay a fine and correct any irregularities before resuming production.

Before exploiting some of the institutional changes in our empirical analysis, we develop a model that generates precise predictions about how AWAs influence the firms' hiring and firing decisions, as well as the decision to employ irregular workers.

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

3 A labor Demand Model of Jobs, Temporary Jobs and AWAs

3.1 The Environment and the Institutions

This section presents a stylized labor demand model in which a firm is characterized by tasks with different probabilities $\lambda \in [0, 1]$ of becoming unproductive.[14] Firms decide which type of contract to offer for each task, with different contracts having varying termination costs. The model considers the possibility of tax evasion by underreporting contracts and explores the impact of labor inspections on firms' use of flexible work arrangements.

Tasks may become unproductive for a variety of adverse technological reasons.[15] Productive tasks generate a homogenous output equal to y. A firm is defined by a finite number $Z/2$ (with $Z \in \mathbb{N}$) of intervals representing tasks.[16] Firms have to decide whether to activate tasks within the ordered set of intervals: $\{[\lambda_1 - \lambda_2], [\lambda_3 - \lambda_4], ...[\lambda_{Z-1}...\lambda_Z]\}$, with $\lambda_j < \lambda_{j+1}$, $j \in Z$.

In other words, we consider a firm that has drawn Z different values from a cumulative distribution function $G(\lambda)$ of tasks with the probability of becoming unproductive below λ, so that $G(1) = 1$ (the corresponding density function is $g(\lambda)$). A task produces output y for the fraction of time $1 - \lambda$, while producing 0 for the rest of the time (normalized to be one). The wage ω paid to each worker for each task is taken as given by the firm.[17] The model is partial equilibrium and we focus only on which type of contract the firm will offer to different λ tasks, although later we are also going to allow firms to hire undeclared workers to evade

[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] λ can also be interpreted as a technological destruction rate: with probability λ the productivity of the job drops to zero.

[16] We assume that Z is even, otherwise one can simply redefine $Z' = 2Z$.

[17] The model can be easily solved with rent sharing.

taxes.[18]

labor regulations allow for three types of regular contract for a given task: open ended, fixed term jobs, and AWAs/voucher (we use the words AWAs 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 a multiple of the wage rate: $F = f\omega$.[19] In what follows, we shall indicate with $J^o(\lambda)$ the value to the firm of a λ task with an open-ended contract. Fixed-term contracts are active for a fraction $1-\rho$ of the time. When a firm opens a fixed-term contract for a task, it commits to pay the worker for an expected duration equal to $1-\rho$, regardless of the specific value of the job of λ. The advantage of a fixed-term contract is that the firm does not pay any firing costs when the expected duration ρ strikes. However, the cost associated with such a contract is that the firm can be forced to pay the worker even if λ strikes and productivity drops to 0. In what follows, $J^{ft}(\lambda)$ indicates the value to the firm of a task regulated by a fixed-term contract. Finally, the firm can open AWAs. AWAs do not have any dismissal cost, but are characterized by an expected duration $1-\rho^v$ where ρ^v is considerably larger than ρ. In practice, it is as if AWAs can be terminated at any time at no cost. In what follows, we shall indicate with $J^{awa}(\lambda)$ the value to the firm of an AWA.

In addition, the labor market is characterized 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. In Section 3.2 we consider the case of tax evasion. There is a fixed cost of opening a given task equal to K, meaning that a firm will open a λ task as long as $J^i(\lambda) \geq K$, where the subscript i stands for open ended contract, fixed term or AWAs.[20]

[18] We are implicitly assuming that workers are ex-ante homogenous and can be employed for any of the tasks.

[19] Setting the tax to be also proportional to the duration of the contract would not change the results.

[20] Assuming different Ks for the different contract does not change any of the following results.

Firms choose the type of contract to offer, conditional on each expected destruction probability λ in one of its intervals:

$$J^*(\lambda) = Max_{\{i=[o,ft,awa]\}} \left\{ J^i(\lambda), K \right\} \quad \forall \ \ \lambda \in \{[\lambda_1 - \lambda_2], ...[\lambda_{Z-1}...\lambda_Z]\} \tag{1}$$

To solve this maximization, we need to specify the expected value of different jobs. The value of an open-ended job is

$$J^o(\lambda) = (1-\lambda)\,(y-\tau-\omega) - \lambda F. \tag{2}$$

The value of the firing tax $F = f\omega$ has the restriction that $f < 1 + \frac{\tau}{\omega}$, so the firing tax for open-ended contracts must be smaller than the regular tax. Otherwise, the firm would be better off keeping an unproductive task and paying the worker. In contrast, the value of a task of λ type under a fixed term contract is

$$J^{ft}(\lambda) = (1-\rho)\,[(1-\lambda)(y-\omega-\tau) - \lambda(\omega+\tau)]\,, \tag{3}$$

where at rate ρ the task is destroyed at no cost. Yet, as argued above- with probability $\lambda(1-\rho)$- the firm is forced to pay the wage until the expected duration.

Finally, the value of an AWA is

$$J^{awa}(\lambda) = (1-\rho^v)\,[(1-\lambda)(y-\omega-\tau)] \tag{4}$$

The maximization problem satisfies the reservation property, since all job values are monotonic and decrease in λ. In other words, maximization is solved by selecting particular values of λ (say λ^{ft} and λ^{awa}) such that each type contract is optimal in any particular subset of the domain of F in $[0,1]$. Furthermore, one can easily show that $J^o(0) > J^{ft}(0) > J^{awa}(0)$.

Furthermore, $J^{awa}(1) = 0 > J^{ft}(1) = -(\omega + \tau) > J^{o}(1) = -F$. The maximization is thus an envelope of three downward sloping lines, and the firm's choice can be described by the two reservation values $\tilde{\lambda}^{ft}$ and $\tilde{\lambda}^{awa}$. The reservation probability can be characterised as the solution to

$$J^{o}(\tilde{\lambda}^{ft}) = J^{ft}(\tilde{\lambda}^{ft}); \qquad \text{and} \qquad J^{ft}(\tilde{\lambda}^{awa}) = J^{awa}(\tilde{\lambda}^{awa}). \tag{5}$$

$\tilde{\lambda}^{ft}$ is the expected duration that makes the firm indifferent between an open-ended job and a fixed-term job. Similarly, $\tilde{\lambda}^{awa}$ makes the firm indifferent between a AWAs and a fixed term job.[21] The intuition of this result is the following. For a given net flow productivity $y-\omega-\tau$, firms have a strong ordering of which task to open according to their expected destruction rate, with open-ended contracts suitable for tasks with a long expected duration and AWAs suitable for tasks with very low duration. Furthermore, AWAs create opportunities for labor demand that would not otherwise be exploited if the AWAs were not there. In other words, AWAs respond to the firm demand for flexibility for jobs with very low expected duration. For simplicity, in what follows, we indicate with $\tilde{y}$ the net flow value of the job so that $\tilde{y} = y - \omega - \rho$.There exists also a maximum λ^{max} that is the solution to $J^{*}(\lambda^{max}) = K$, and the firm does not activate tasks for any $\lambda > \lambda^{max}$. If $\lambda_Z < \lambda^{max}$, the total employment of the firm is therefore $n = \sum_{j=2}^{Z} F(\lambda_j) - F(\lambda_{j-1})$, where j is even. In contrast, if $\lambda_Z > \lambda^{max}$, the total employment in the company is thus $n = \sum_{j=2}^{T+1} F(\lambda_j) - F(\lambda_{j-1})$, where j is even and λ_T is the highest value of λ in the firm set, so that $\lambda_T \leq \lambda^{max}$.

[21] The threshold λs are:

$$\begin{cases} \tilde{\lambda}^{ft} & = \frac{\rho\tilde{y}}{\rho\tilde{y}+(f\omega-(1-\rho)(\omega+\tau)} \\ \tilde{\lambda}^{awa} & = \frac{(\rho^{v}-\rho)\tilde{y}}{(\rho^{v}-\rho)\tilde{y}+(1-\rho)(\omega+\tau))} \end{cases} \tag{6}$$

Note that existence of two thresholds- and thus two fixed term contracts- require that the duration of AWAs is sufficiently short, or that

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

3.2 Shadow Employment and the Misuse of AWA

We now introduce the possibility of evading taxes by under-reporting contracts associated to specific tasks. A shadow task allows firms to avoid paying the tax τ. In terms of the type of contract, we talk about a general λ task, and let $J^{i,s}(\lambda)$ be the value of a representative λ task that is employed with a shadow job, or irregular worker, where i refers to the 3 types of job contracts. For simplicity, we assume that firms are all identical with respect to their tendency to go shadow, and we do not add any further dimension of heterogeneity across firms.

In addition, γ is the probability of inspection and $C(\lambda)$ is the fine imposed on the firm with undeclared work upon inspection. The main assumption we make is that $C'(\lambda) < 0$. According to common practice, inspectors will charge a higher fine to workers who appear to be in a longer-lasting employment relationship (and therefore have lower λ). The decision to go shadow is simply

$$J^{i,s}(\lambda) = (1-\gamma)(J^i(\lambda) + \tau) + \gamma(J^i(\lambda) - C(\lambda)) > J^i(\lambda) \qquad i = \{O; F; AWA\}$$

which implies the standard conditions found in most of the shadow employment literature, namely that $(1-\gamma)\tau > \gamma C(\lambda)$ so that going shadow is optimal if the tax evaded is larger than expected fine. At the margin, this implies that if the expected duration $1/\lambda$ is sufficiently low (or the adverse shock sufficiently high), the firm will operate the task with a shadow worker or unreported employment[22]

How does the decision to go shadow change, when vouchers become available? Let us start with the case where AWAs can be activated on the spot. When labor inspectors appear, firms have the option to declare that the task is covered by a voucher, which means that they can use vouchers as an insurance mechanism. We define these jobs as "gray:" they

[22]See Boeri and Garibaldi (2005) for a matching model with shadow employment.

are legitimate but hidden in black as long as the inspector does not show up. Let $\tilde{J}^{i,s}(\lambda)$ be the value of an irregular task that has the option to activate the voucher conditional on inspection, or the value of a shadow job that has the option to misuse vouchers.

Formally, the existence of vouchers adds an additional choice, generating an option value. The decision to go shadow with a misuse of vouchers corresponds to the following case:

$$\tilde{J}^{i,s}(\lambda) = (1-\gamma)(J^i(\lambda)+\tau) + \gamma \underbrace{\left(Max[J^i(\lambda) - C(\lambda); J^{awa}(\lambda)]\right)}_{option\ to\ misuse\ vouchers} > J^i(\lambda). \tag{7}$$

The implicit assumption behind equation 7 is that it is impossible conditional on an inspection to offer a regular open ended or regular fixed term job to workers who have been hidden behind a voucher. The previous maximization is certainly satisfied for AWA, while it is not obvious in the case of open-ended and fixed-term workers. If, on the contrary, AWAs cannot be activated on the spot, firms that choose to misuse vouchers need to always buy a voucher to hide a worker. In this case, the corresponding value would be equal to the previous one minus the minimum cost of the vouchers ($\epsilon\gamma$), where ϵ is the cost of vouchers.

The general firm problem with both shadow employment and the option to misuse is thus

$$J^*(\lambda) = Max_{\{i=[o,f,awa]\}} \left\{ J^i(\lambda), J^{s,i}(\lambda), \tilde{J}^{s,i}(\lambda), K \right\} \quad \forall \ \lambda \in \{[\lambda_1 - \lambda_2], ...[\lambda_{Z-1}...\lambda_Z]\} \tag{8}$$

The existence of various thresholds suggest that the model is fairly flexible. The solution can be obtained in two steps: first, for each type of contract, it is possible to solve the option value problem. In the second step, the company chooses the best contract.

3.3 Inspection, Misuse of AWA, and Their Abolition

The main results from the previous section are that i) regular employment is more likely among open ended contracts and ii) firms are going to misuse voucher to hide undeclared work (by definition of equation 8). To solve an explicit example of the model, we assume a linear penalty cost $C(\lambda) = C_0 - c_1\lambda$, with both $C_0 > 0$ and $0 \leq c_1 \leq 1$. With respect to the distribution $G(\lambda)$, the implication of the model outlined in the previous section is that is that within the distribution $G(.)$ there is a sub set of task covered by fixed-term contracts and a sub set covered by AWA. By adding the option to go shadow, this section shows that unreported employment exactly in the sub set of tasks covered by fixed-term contracts and AWA.

In addThese are the tasks in which unreported employment is more likely to emerge in real-life labor markets.

The example we carry out is described in Figure 2. The firm potentially operates in 4 different intervals (i.e. $Z = 8$ and $Z/2 = 4$ intervals are indicated in yellow in the Figure) and the solution to the labor demand problem is to use all types of contracts.[23] In each subset of tasks, the optimal contract chosen is derived from the maximum value of the linear value function corresponding to that particular contract. However, the firm can also go shadow in some of its active tasks. The top panel of Figure 2 thus simulates a labor demand problem of a firm that optimally hires regular open-ended and temporary workers, while it hires as shadow workers those who work on tasks above λ_5. In the second panel, right after the inspection, all the shadow workers are officially employed using vouchers. The firm pretends that these workers are regularly hired under AWA schemes. When AWA can be activated on the spot, the prediction is that on average we should expect an increase in AWA on the day of inspection, which is exactly the scenario described in the second panel.

We therefore have that if

$$J^{AWA}(\lambda) > J^{i}(\lambda) - C(\lambda), \tag{9}$$

[23]The parameters used in the example are $y = 1.4; \omega = .5; f = 1.7; \tau = 0.2$. The arrival rates of the contracts are $\delta_t = 0.18$ and $\delta_v = \delta_t + 0.6$. The entry cost $K = 0.12$, the inspection arrival rate is $\gamma = 0.06$, while the fee function is $C = 0.4y + 0.45\lambda$.

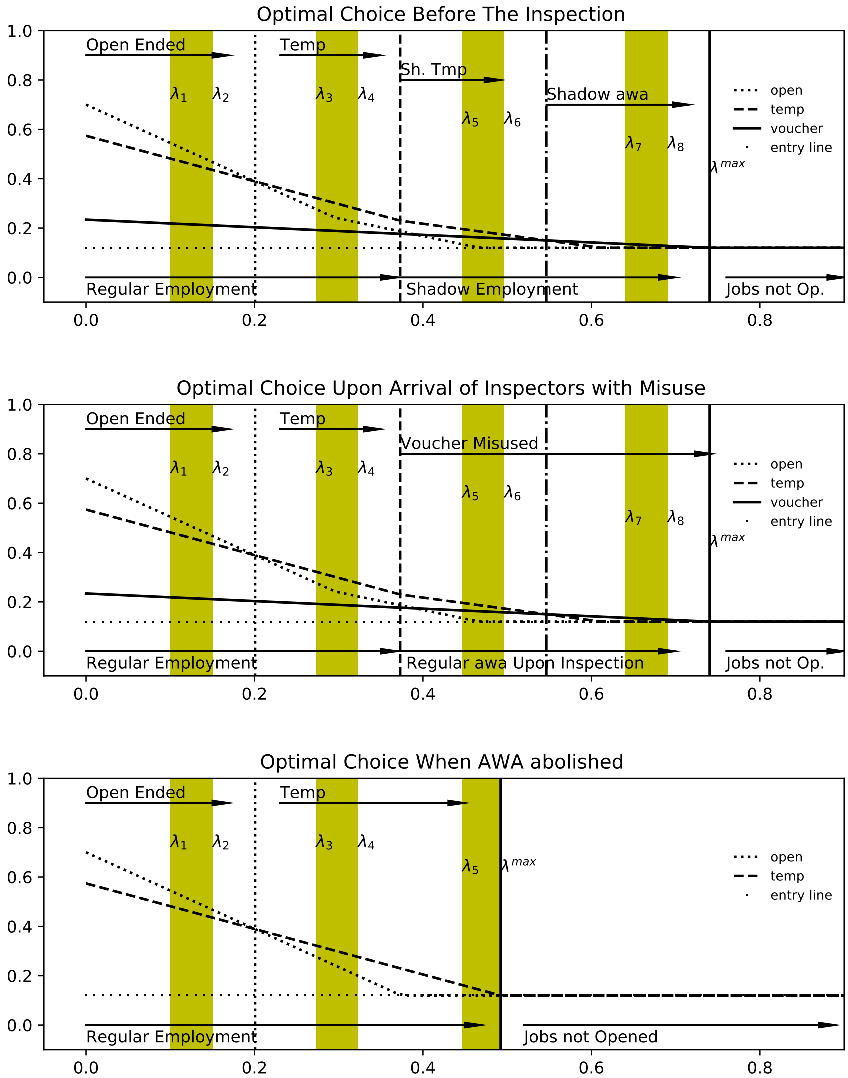


Figure 2: Optimal labor Demand when AWA can be misused, and when they are outlawed.

Notes: The top panel simulates a labor demand problem of a firm that optimal hires regularly open ended and temporary workers, while it hires as shadow workers some temporary workers as AWA. The shaded area in the top panel corresponds to the intervals for which the firm has active tasks. In the middle panel, shadow employment is covered using AWA when the inspection starts. The bottom panel AWA are abolished and the firm operates only in three intervals, with open-ended and temporary contracts.

firms activate vouchers upon inspection, which is the case in the second panel of Figure 2. *The misuse of vouchers upon inspection represents the first testable implication of our model.*

Let us assume that for a subset of firms, equation 9 is satisfied. This, in turn, implies that the expected tax evaded is larger than the expected costs of musing vouchers, or more formally

$$\underbrace{(1-\gamma)\tau}_{\text{expected tax evaded}} > \underbrace{\gamma(J^i(\lambda) - J^{AWA}(\lambda))}_{\text{expected cost of misusing voucher}} \tag{10}$$

Yet, from Equation 9 we know that $J^i(\lambda) - J^{AWA}(\lambda) < C$, thus *the possibility of misusing vouchers makes it more profitable to exercise the option to go shadow.* This prediction represents the scenario highlighted in the second panel of Figure 2 and it is our second empirical implications. Indeed, in the Figure, all the tasks above 0.38 are employed under a shadow employment.

The lower panel of Figure 2 predicts firm's labor demand when AWA are abolished. Shadow employment goes down, in this particular case all the way to zero. At the same time, after the abolition of vouchers, firms increase regular employment through a greater use of fixed-term employment. This is our third testable implication.

In summary, the three results that follow from equation 9 and the discussion in this section are as follows.

1. Some firms may misuse vouchers and activate them upon inspection.

2. The amount of shadow work (through the misuse of vouchers) increases.

3. Regular employment increases at the firm level if vouchers are prohibited.

Next, we test the predictions of our model.

4 Data

The study uses data from the INPS archives, including firm-level employment data, daily firm-level voucher purchases, and labor inspection records. The analysis focuses on firms that have been inspected and have used vouchers, constructing daily-level data on voucher use around labor inspections

In particular, we start with the universe of firm-level employment data for the years 2014 to 2017. As shown by the timeline (Figure 1) and the time series of vouchers used (Figure 3) these years represent the peak years, where vouchers were almost completely liberalized.[24] Furthermore, we merge firm-level economic data with both the universe of labor inspection and the universe of vouchers used by firms.[25]

The INPS institute performs labor inspections to detect full or partial evasion of social security contributions. We have information on the day the inspection started and on the outcome of the inspection (whether a fine was levied and its amount). For each firm, we know how many vouchers have been used each day. Table 1 documents the entire data selection based on monthly summary statistics for the period January 2015 to September 2016.

We start with the universe of 1.8 million firms, who employ an average of 7.2 workers (column 1, Full). Most of these 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.

The average monthly number of vouchers is large and equal to 97.6 (15.7 vouchers per full-time equivalent worker). The third column (Inspected) refers to inspected firms, which represent only 1.58% of firms. Since larger firms are more likely to be inspected, the average

[24] Similar data have already been used within the VisitInps program of the Italian Social Security institute (INPS).

[25] 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.

workforce increases to 43.5 workers. In addition, inspected firms use a large number of vouchers, on average 300 vouchers per month, or 7.74 vouchers per full-time equivalent worker.

If each voucher covered one hour of work, the vouchers would represent only a small share of labor, but we are going to see that firms use vouchers to hide additional undeclared work hours.

When we focus on firms that have been inspected and use at least one voucher in the period, we are left with 3472 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 270 (9.2 for the full equivalent worker). This set of firms that use vouchers and are being inspected represents our main reference sample for the first part of the analysis. This sample provides information on how firms using vouchers, which represent almost 25% of the universe of Italian firms, behave upon inspection. We will comment later on the last two columns, where we divide the 3472 firms into those who presumably misused vouchers ("jump up") and those who did not ('jump down").

Furthermore, since we know exactly the day vouchers are used and inspections take place, we can construct daily-level data on the use of vouchers around labor inspections. For subsequent analyses, we use monthly data, as the number of workers within each firm, for each type of labor contract, is recorded every month.

5 Empirical Evidence

First, we discuss the identification, which is based on the random timing of labor inspections (Section 5.1). Then we test whether firms are more likely to use vouchers when inspectors appear at their premises (Section 5.2). Finally, we analyze what happens to different types of regular and irregular work vouchers are abolished (Sections 5.3 and 5.4).

5.1 Identification

Inspections are rare and their timing is unpredictable, allowing for a comparison of firms' behavior 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 being inspected in a given day. Although some firms are more likely to be inspected than others–for example, larger firms–for inspections to be effective, the timing of inspections is unpredictable. Thus, from a firm's perspective, the first day that labor inspectors enter the firm's premises is as good as random.

Given this randomness that we demonstrate in Section 5.2.1, our model suggests a fairly simple test to determine whether an inspected firm is using vouchers to hide undeclared work: when faced with labor inspectors, firms employing undeclared workers should "exercise the option" to use vouchers. Thus, we compare the daily use of vouchers just before and after an inspection (the "treatment"), following the firm's behavior several days before and after the inspection. We use all companies that have been inspected at least once between 2014 and 2017 and have used at least one voucher throughout the period.

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 labor inspection. We use as a balance test the joint F-test that 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, it is straightforward to analyze the AWAs used by the firm j between 180 days before and 90 days after an inspection that occurs on day t.[26] We perform two analyses, before and after October 2016, which is when the firms had to

[26]Labor judges have deemed that 90 days represent the reasonable duration of labor inspections Del Vecchio (2019).

inform the Social Security Administration an hour in advance before using a voucher. Given that a single voucher would be sufficient to avoid the fine, 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]

$$DV_{j,\tau} = \sum_{k=-90}^{90} \beta_k D_{\tau+k} + f(t) + \epsilon_{j,\tau} \; . \tag{11}$$

$D_{\tau+k}$ is a dummy variable equal to one for event day $\tau + k$ and zero otherwise. Additionally, since the time series of AWAs is far from stationary (see Figure 3) it is important to control for 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 are collinear. Fortunately, when the timing of the event is random, and we will provide evidence of it, the timing of treatment is orthogonal to the characteristics of the firms (or their time-invariant intercept (see Borusyak et al., mimeo). Our strategy is similar to the identification used in Parker et al. (2013), which exploits the randomized 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-

[27]We find comparable results when using the daily number of vouchers used by the firm, though the extensive margin appears to be driving this effect. This suggests that most (declared) firms have only a few undeclared workers.

[28]Given that the 90 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.

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 nonlinear 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.

Yet, it is worth mentioning that without fixed firm 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, using both firm effects and calendar-time effects captures most of the variation in the timing of inspection.

We are going to discuss later how we deal with these potential biases. Here, it is important to mention that with the random timing of inspections, we can use the main result in 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 to differ, disregarding any collinearity issues, the Difference-In-Differences identify the average causal effect.[30]

In order to spot misbehaving firms, we exploit that treatment effects appear to be stable over time. With firm-specific constant treatment effects, we can collapse the coefficients and use the entire pre-inspection period as the baseline:

$$DV_{j,\tau} = \beta_j D_{\tau \geq 0} + f(t) + \epsilon_{j,\tau} \ , \tag{12}$$

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

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

vouchers.

Based on the estimated $\hat{\beta}_j$s, we identify misbehaving firms and highlight how they respond to the introduction of the SMS[31] requirement and the abolition of vouchers. In particular, we combine the results from the previous model with the October 2016 SMS requirement and the March 2017 abolition of vouchers, separating firms with positive changes from the rest: $\widehat{M_j^\eta} = 1\{\hat{\beta}_j > \eta\}$, where η represents the cutoff above which a firm misbehaves: in the baseline analysis $\eta = 0$. Later we test whether the results are robust to the choice of more stringent cutoffs, $\eta > 0$.

In other words, we use behavioral changes driven by the inspections to identify firms that are likely to misuse vouchers: for each inspected firm, we compute their average use of vouchers before and after the inspections and classify firms into those who on average increase their use and those who do not. Our definition of misbehaving firms is subject to misclassification, both of type I and II, which biases the estimates towards 0. We will take advantage of the possibility of changing our definition of misbehaving firms.

Our model predicts that firms that are likely to misuse AWAs ($M_j = 1$) would start buying more vouchers in October and, a few months later, when vouchers are abolished, would fall back into signing contracts that allow for some "gray" work, such as part-time or fixed-term contracts, revert back to hiding the entire work relationship, or abandon the low-productivity task.

Empirical models are simple difference-in-differences, before and after October 2016 or March 2017 between firms that presumably misused AWAs and those that did not. We are implicitly assuming that misbehavior is time-invariant, and any deviation would lead to downward biased estimates

Since all firms share the same event date, we are not in a staggered design and do not have to worry about biases of the dynamic treatment effects (see Sun and Abraham, 2020). To

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

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, allowing for some anticipation effect).

The outcomes available at the monthly level (m) are i) total number of vouchers used, the ii) log of 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:

$$Y^m_{j,t} = \sum_{k \neq -2} \gamma_k \widehat{M}^0_j \times D_{\tau(t)+k} + \mu_j + \mu_t + \varepsilon_{j,t} \ , \tag{13}$$

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:

$$Y^m_{j,t} = \gamma \widehat{M}^\eta_j \times D_t + \mu_j + \mu_t + \varepsilon_{j,t} \ , \tag{14}$$

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 $\frac{\widehat{\gamma}(\eta)}{1-p(\eta)-q(\eta)}$ converges in probability to γ. This implies that with constant treatment effects the η^* that maximizes $\widehat{\gamma}(\eta)$, minimizes the misclassification bias. Later, we discuss whether η can tell us something about the fraction of firms that misuses flexible work arrangements.

In our final analysis, we look at how the introduction of the SMS requirement, as well as the abolition, has influenced under-reporting. Since under-reporting is not directly observable, we need to rely on labor inspections. Yet, since only a handful of firms are inspected more than once in our data, we cannot compare misbehaving with non-misbehaving firms, but need to rely on coarser definition of treatment. In particular, we compare firms based on whether they have ever used a voucher before the SMS requirement.

An observation will be a labor inspection between January 1, 2016 and December 31, 2017, which will produce 20,819 observations. The treated firms are those that have used vouchers in the pre-SMS period, 4,269 observations. We use as outcome variable the evaded contribution to measure underreporting and define three main treatment periods: Pre-SMS requirement (May 1 to October 16, 2016), post-SMS requirement (17 October 2016 to March 17, 2017), and post-abolition (March 18, 2017 to December 31, 2017).

5.2 The Misuse of AWAs: Evidence from Labor Inspections

5.2.1 Random Timing of labor Inspections

We start by testing whether the timing of the labor inspection is predictable. Appendix Figure A1 shows the distribution of the day of the year inspections take place. Inspections occur throughout the year and are fairly uniformly distributed, although 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 labor inspectors to have different targets in different years. Only three variables are 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. All differences, with the exception of the residual sector "Other," become less than 3 days.

In other words, the exact day that inspections take place appears to be unpredictable.

5.2.2 Vouchers and labor Inspections

Given the 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 11 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 days between 90 and 180 before 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 pretrends in the use of vouchers prior to the inspection, ii) no major anticipation effects, and iii) fairly stable treatment effects.

Appendix Figure A2 shows that without calendar time controls the probability of using at least one voucher grows over time almost linearly (right panel), but that adding a simple linear time trend is enough to center the pre-period around zero (left panel).

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 are large effects. Furthermore, the figure shows that these effects persist for at least 90 days. This is likely due to the fact that inspections last several weeks, and we cannot exclude that misbehaving firms may update their posterior probability to be inspected and start using vouchers more frequently.

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

$-120 \leq \tau \leq -30$ and estimate

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

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 Figure 8 is centered around zero and is far from the vertical line, which corresponds to the estimated $\widehat{\beta}_0 = 0.88$. The sampling noise is unlikely to have generated such a large change in behavior.

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

$$DV_{j,\tau} = \beta D_{\tau\geq 0} + f(t) + \epsilon_{j,\tau} \ . \tag{16}$$

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. This may depend on the fact that in the construction sector work injuries are so common that firms prefer to buy a voucher per casual worker per day to insure them against work-related accidents, regardless of being inspected.

In Table 3 we perform additional heterogeneity regressions. Columns 1 to 4 show that the jump is fairly constant across Italian regions, although it is slightly larger in the more productive north than in the south, with the center of Italy being in between the other two regions. 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). Finally, the last column 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 labor 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.

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 A3, 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 A1 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 fixed effects for day 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. Firm fixed effects lower the standard errors, leading to very precisely estimated coefficients.

5.2.3 The 60-minutes SMS Requirement

Next we analyze the October 2016 SMS reform, which introduced the 60-minute messaging requirement and thus broke the possibility of on-the-spot insurance. Firms would still be able to exercise the insurance option on a daily level, buying at least one voucher per worker, but this should not generate a jump on the day of the inspection.

Figure 6 shows, indeed, that once SMS was introduced, the jump disappears.[32] 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.

To test whether the total number of vouchers increased after October 2016, we use monthly data and define as “misusing” companies those who before the SMS requirement and upon inspection were more likely to use vouchers. Figure 7 shows the event-study differences between “misusing” and “non-misusing” firms, where 0 refers to the timing of the reform. The estimates are noisy, possibly because the reform may have also pushed some misbehaving firm to stop using vouchers altogether, but firms appear to use 10-15 more vouchers per month, which represents a 10% increase compared to baseline, and there is evidence of some anticipation, possibly because the introduction of the October SMS requirement was known several months before implementation.[33]

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

5.3.1 The Effect of Vouchers on Regular Employment

We have shown that AWAs can be used to hide undeclared work, even when informational requirements are designed to avoid such behavior (the SMS). 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 gray).[34]

[32]Less than five coefficients out of 90 are positive and significant, including the day of the start of the inspection. Given the shorter time period that is available we use -60 to -30 event days as a comparison period.

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

[34]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 lower bound for declared and undeclared job opportunities.

Our theoretical model predicts that in the event that AWAs become unavailable, firms should hire some of the "gray" workers using the next most flexible work arrangement (see Figure 2).

The next most flexible work arrangement is arguably the combination of temporary and part-time labor market contracts.

Using monthly data, we investigate over time differences between "misusing" and "non-misusing" firms in a event-study difference-in-difference approach. Figure 9 shows that temporary part-time contracts increase by about 1, representing a 50% increase compared to the pre-abolition average. The difference between "misusing" and "non-misusing"firms in the total number of temporary part-time employees is fairly flat in the months leading to the abolition of vouchers in March 2017 and then increases by approximately one additional worker.

These workers also drive the results for the total number of workers (see Figure 10), since this change is only slightly above 1 (the average number of workers is around 40). Figure 11 separates the effects for the two dimensions: temporary vs. permanent (top panel) and part-time vs. full-time (bottom panel). The greatest 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 labor contracts.

Regarding the total wage bill, when we include the cost of AWAs, there is no evidence of significant changes (see Appendix Figure A4). This suggests that there are no major changes in the job opportunities declared, possibly driven by rule-of-thumb cost calculations.

5.3.2 Misclassification Error

Based on the evidence on the closest substitute contract for AWAs, we use temporary and temporary part-time contracts to assess the importance of misclassification errors. Misclassification is driven by the way we identify "misusing" firms in the baseline event-study regression. When defining $\widehat{M}_j^{\eta} = 1\{\hat{\beta}_j > \eta\}$, with $\eta \geq 0$, type I and type II errors depend on η. Thus, we re-estimate equation 14 comparing "misusing" and "non-misusing" firms, varying ηs between 0, the baseline case, and 7.5%, which corresponds to, respectively, 50 and 8% of misusers. Intuitively, when we increase η we impose a stricter requirement to be identified as a firm that is misusing vouchers. This reduces the likelihood of classification as misusing a firm that is correctly using vouchers (Type II error), but increases the misclassification probability of misusing firms (Type I error).

Figure 12 shows that regardless of whether we use temporary contracts or temporary and part-time workers, the optimal η that minimizes the sum of Type I and II probabilities is close to 0.03, corresponding to a fraction of misbehaving firms of about 18 to 19%.[35] When compared to the coefficients estimated with $\eta = 0$, the coefficients increase by about 60%, which implies that the relative changes in the number of temporary and temporary part-time workers increase from 50% to more than 75%.

5.4 The Effect of Vouchers on Evasion

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

We can only compare the amount of social security contributions evaded for inspected

[35] It is comforting that the optimal η also corresponds to one of the two peaks in the R-squared that are close to each other in size.

firms that have or have not used vouchers between January and October 2016, a clearly less appealing type of comparison. Separating the inspected firms according to whether they have used at least one voucher is suboptimal, we have just seen that a minority of firms use voucher 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. Table 4 includes the difference-in-difference estimates of a model where time periods are interacted with the treatment variable, that is, companies using vouchers, and time periods (Pre-SMS, from May 1 to October 16, 2016; Post-SMS, from October 17, 2016 to March 17, 2017; Post-Abolition, from March 18 to December 31, 2017). Four different specifications are reported, where column (1) does not include any additional control variables to the baseline difference-in-difference, while from column (2) to (4) several sets of controls are added.

It emerges that, compared to control firms, firms that used vouchers appeared to evade more when vouchers became unavailable, while at the height of the use of vouchers (interactions with Pre and Post-SMS) evasion appeared on average lower. The results change little when we control for the share of part-time workers, fixed-term workers, its interaction, average wages, and a cubic monthly trend.

6 Conclusions

In response to the increasing demand from firms for flexible work contracts, legislators around the world are devising labor contracts that allow firms to hire workers for just a little amount of work, i.e., on-call work, zero-hour work, labor 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 labor contracts may complicate the job of labor inspectors whose task is to uncover undeclared work. We show that upon random inspections, firms use alternative work arrangements to hide undeclared work. This counteracts the expected reduction in undeclared work resulting from the reduction in hiring and firing costs.

There is also evidence that small changes to the bureaucratic requirements of alternative work arrangements can partially close these loopholes. When employers were forced to signal such arrangements in advance, labor inspectors would have enough time to identify undeclared workers. This forced misbehaving firms to buy at least one voucher every day to avoid being fined by the inspectors. We estimate that 18% of firms with AWAs, used AWAs to hide undeclared work, and that these firms hired more fixed-term and part-time workers when AWA were banned. This evidence and the evidence that the total wage bill stayed constant suggest that firms did not lower the demand for declared labor when AWAs became unavailable.

In general, our results suggest that in countries with large shares of undeclared work, where alternative work arrangements introduce flexible work schedules, shadow employment may get worse. Moreover, firms who use vouchers to hide undeclared work tend to use vouchers mainly when inspected, thus restricting the total number of vouchers used by workers or firms to be below some threshold, which has been Italy's primary restriction, is unlikely to successfully curb misbehavior. Rules that restrict the use of AWA 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.

Bruno Anastasia, Saverio Bombelli, and Stefania Maschio. Il lavoro accessorio dal 2008 al 2015. profili dei lavoratori e dei committenti. *Collana WorkInps Papers*, (2), 2016.

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.

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. Shadow sorting. In *NBER International Seminar on Macroeconomics*, pages 125–163. University of Chicago Press, 2005.

Tito Boeri and Pietro Garibaldi. Temporary Employment in Markets with Frictions. *Journal of Economic Literature, forthcoming*, 62(4), 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. Technical report, 2021 mimeo.

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.

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

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.

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.

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.

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.

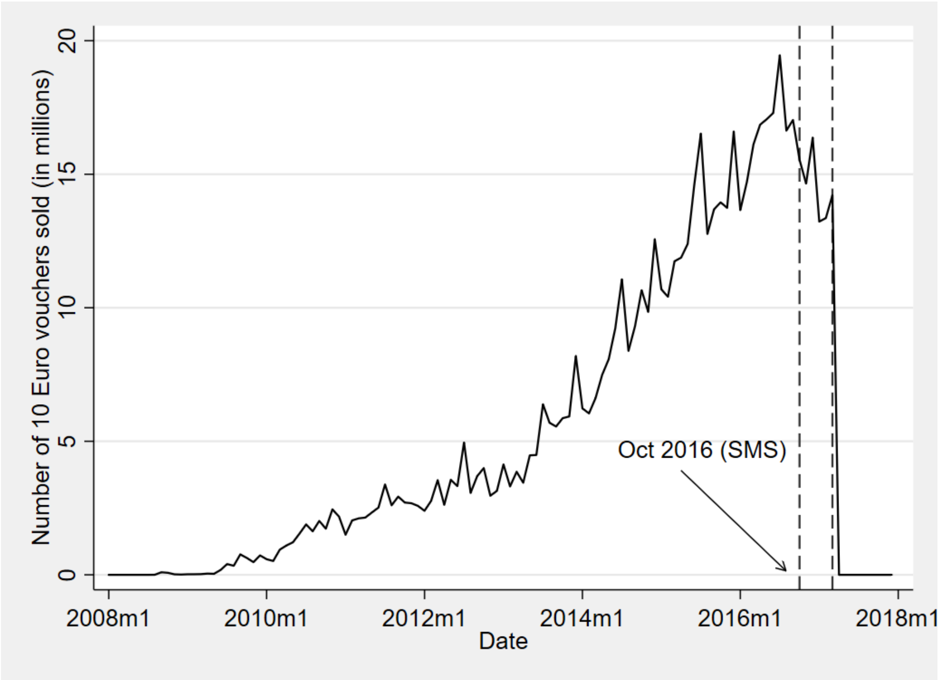


Figure 3: Vouchers Sold

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

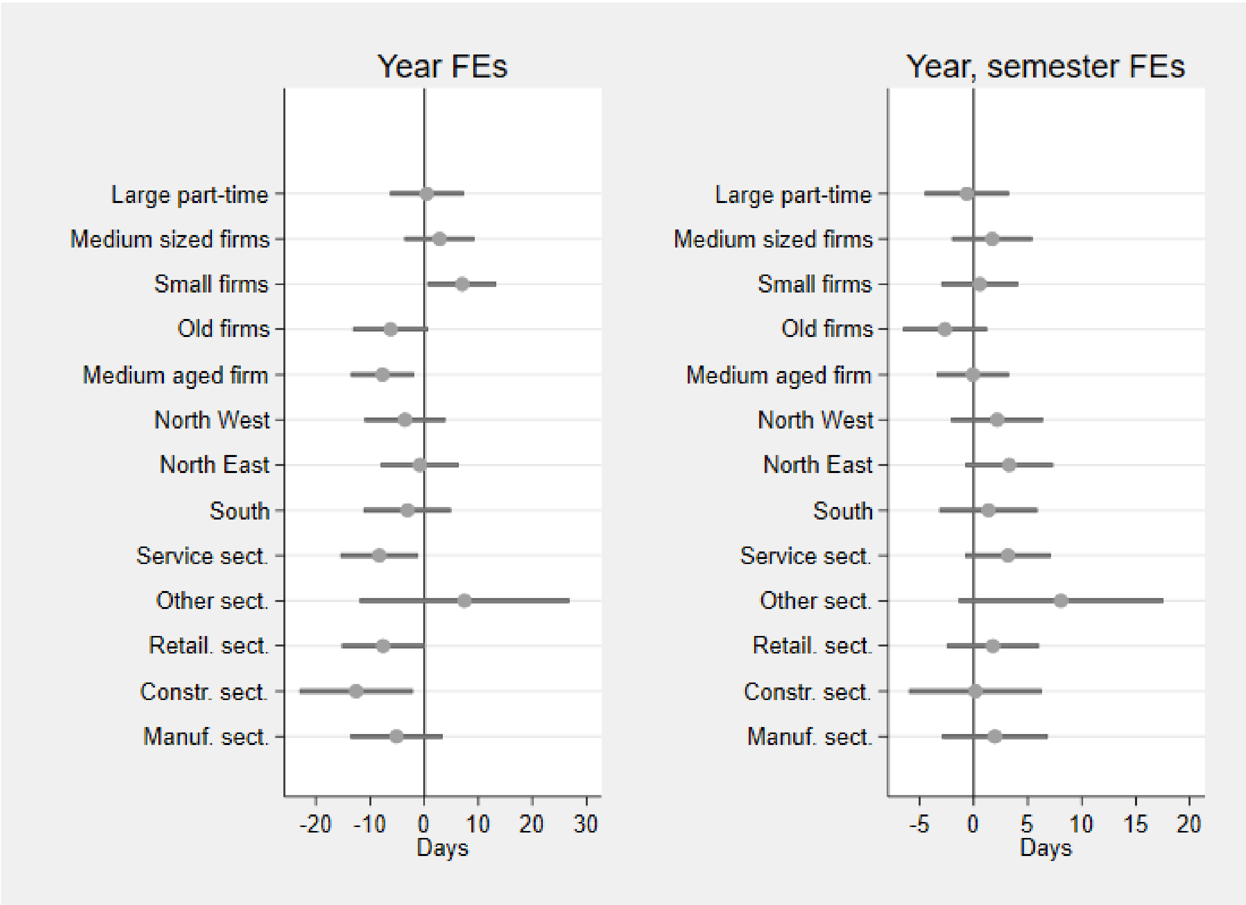


Figure 4: Balance Test

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%.

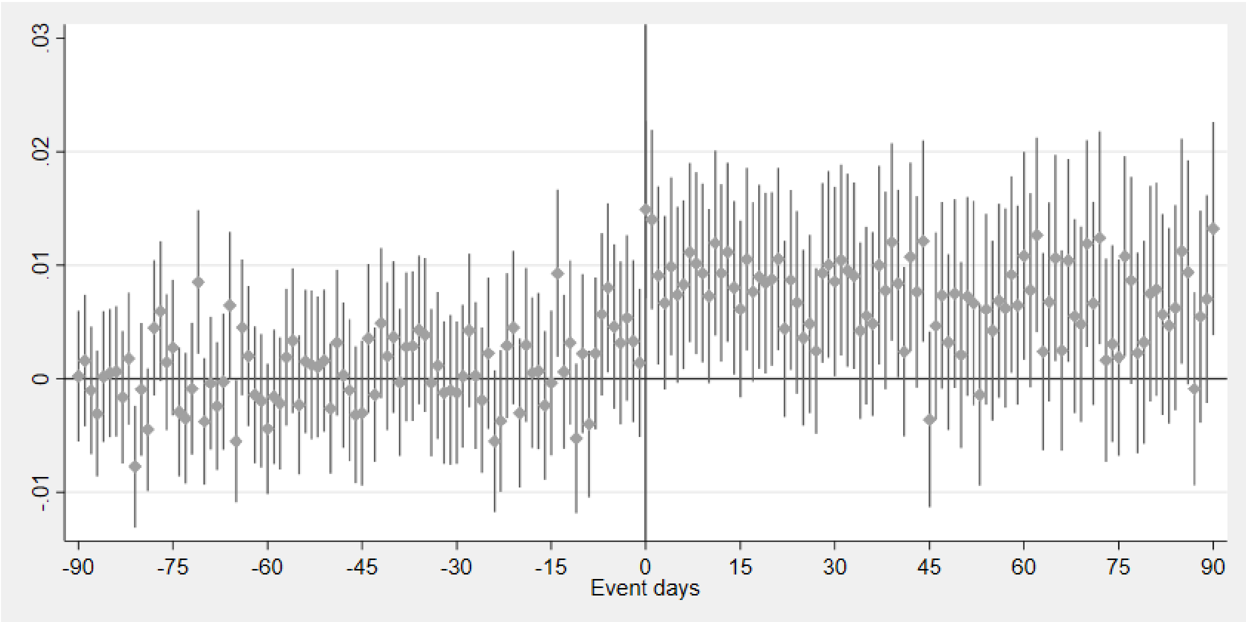


Figure 5: Event Study pre-SMS

Notes: The figure plots event study coefficients, where the event is a labor 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.

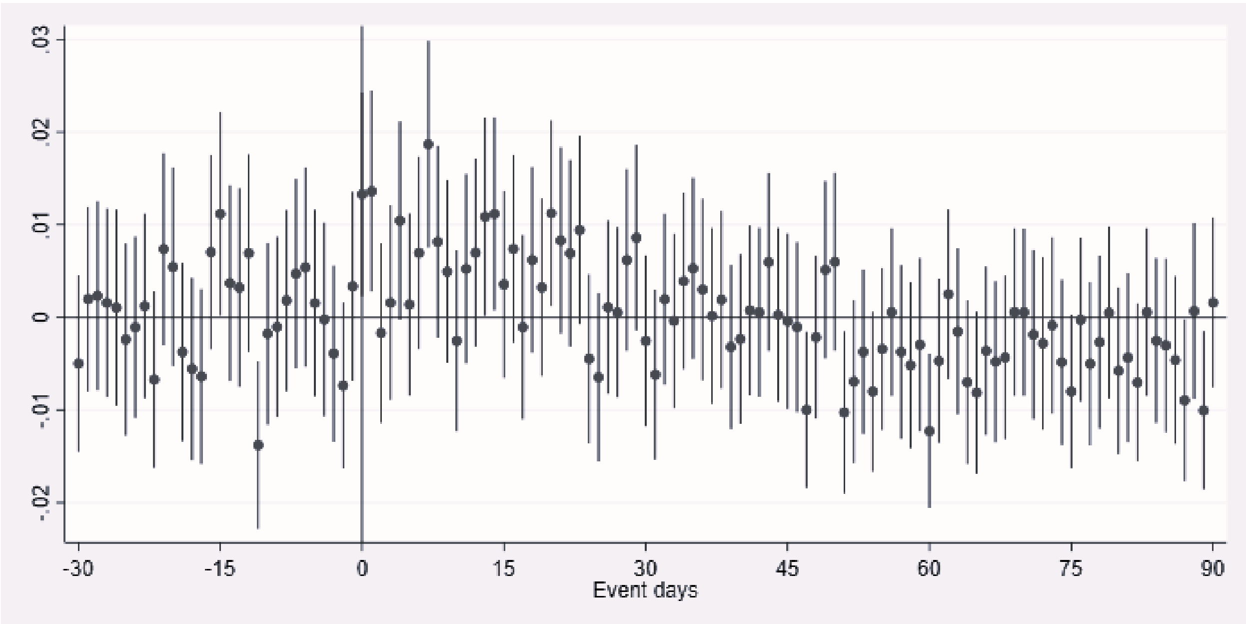


Figure 6: Event Study post-SMS

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

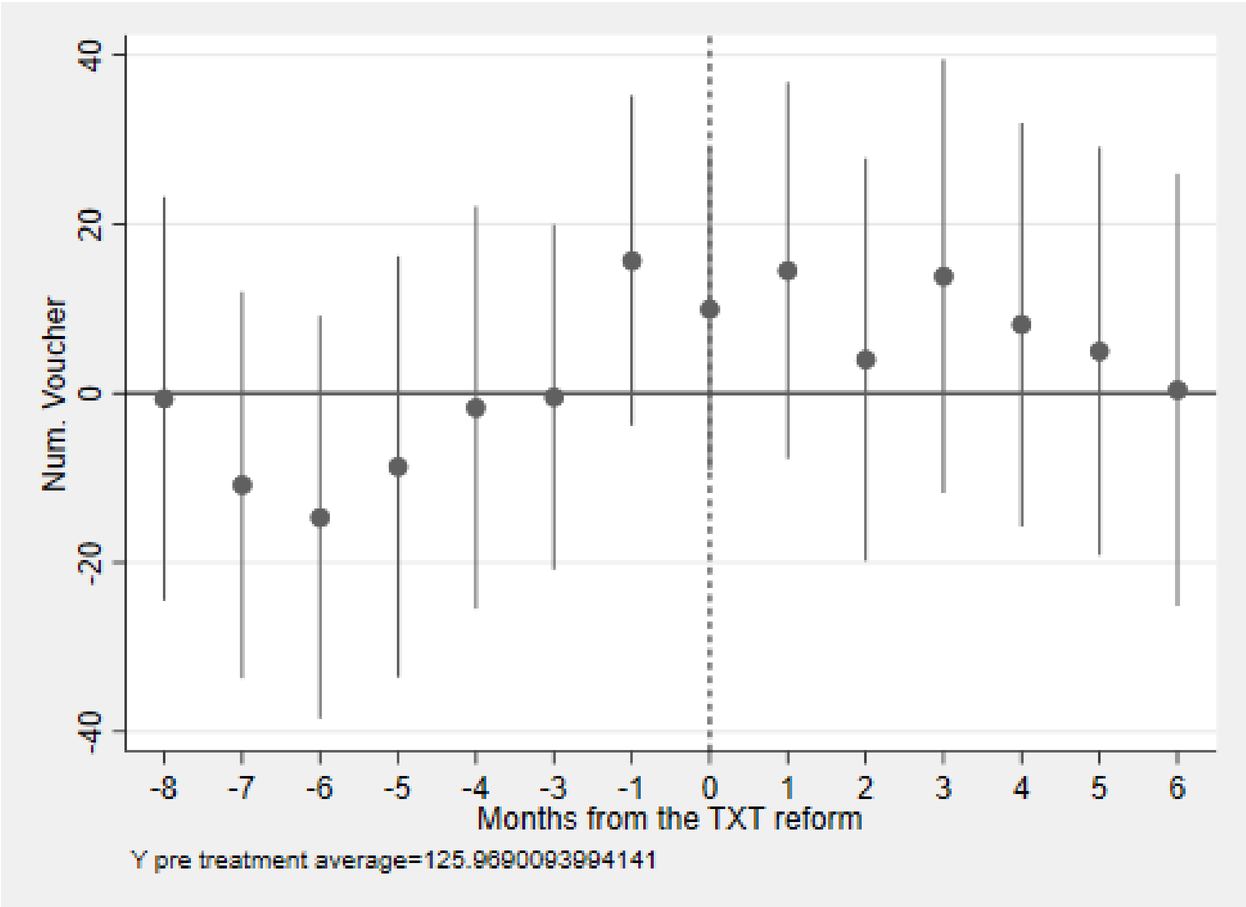


Figure 7: Event Study post-SMS

Notes: The figure plots differences in the number of vouchers used at firms that on average "mis-used" 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.

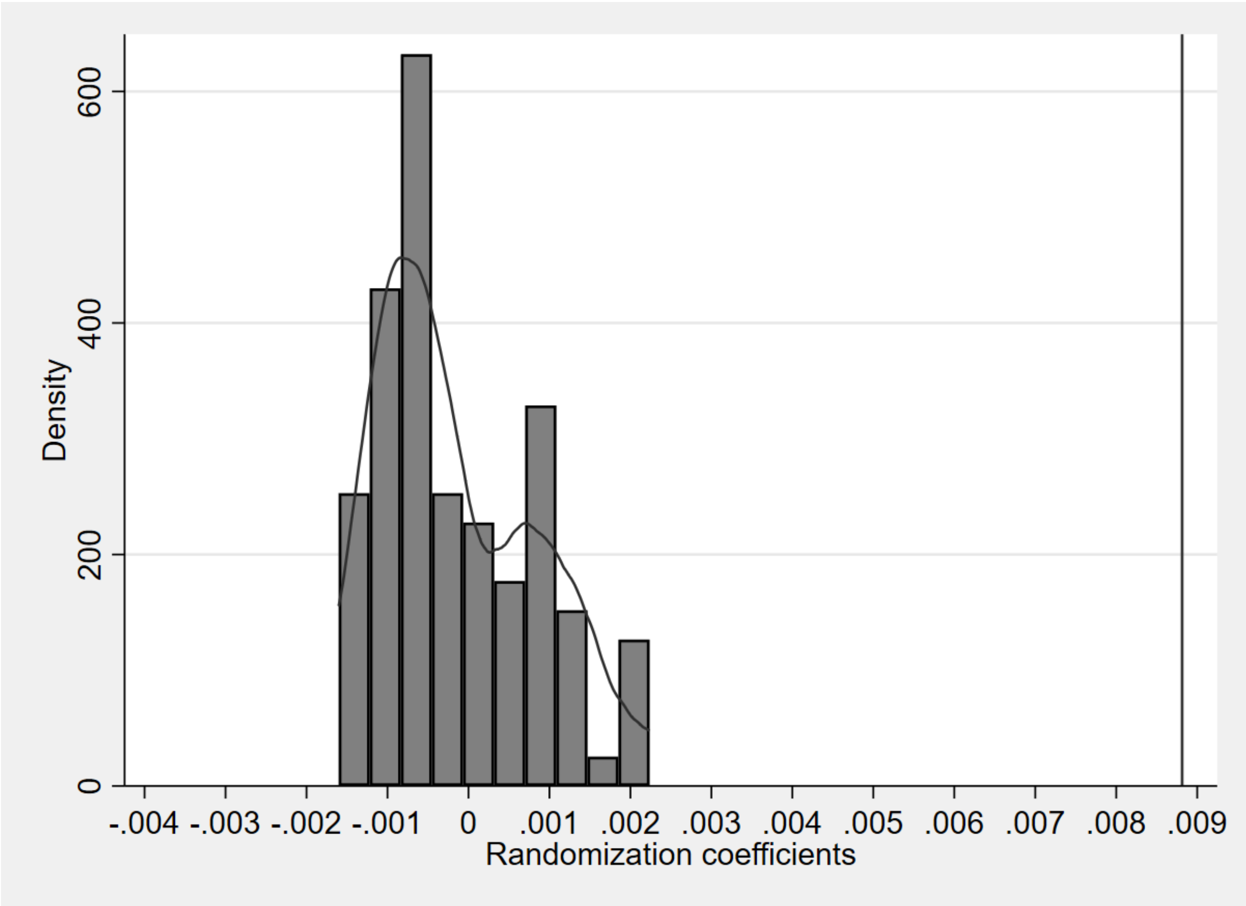


Figure 8: Randomization Test

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 labor inspection is correct ($t = 0$) and the full sample is used ($t \lesseqgtr 0$).

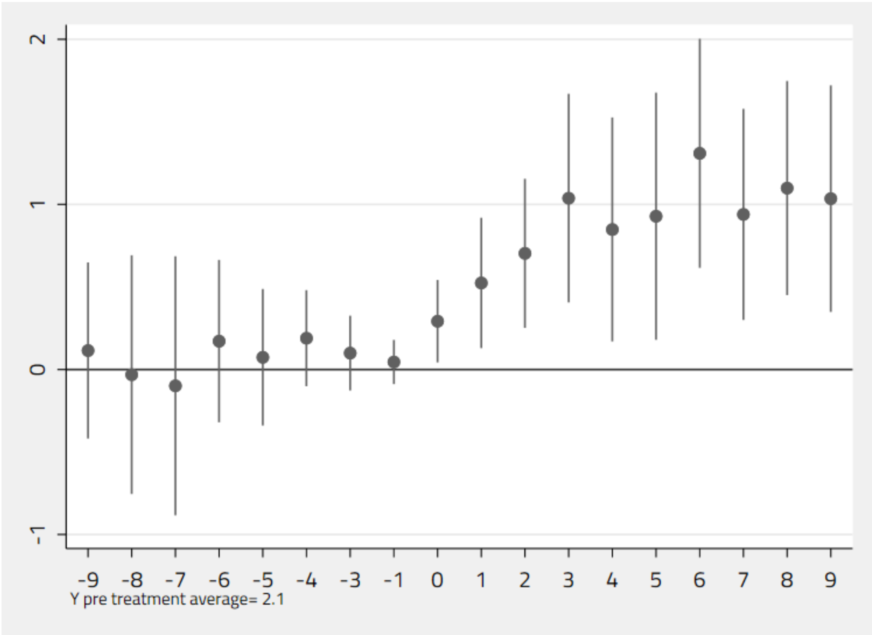


Figure 9: Part-time and Temporary Contracts Around the Abolition of Vouchers

Notes: The figure plots differences in the total number of part-time and temporary workers employed at firms that on average "mis-used" 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.

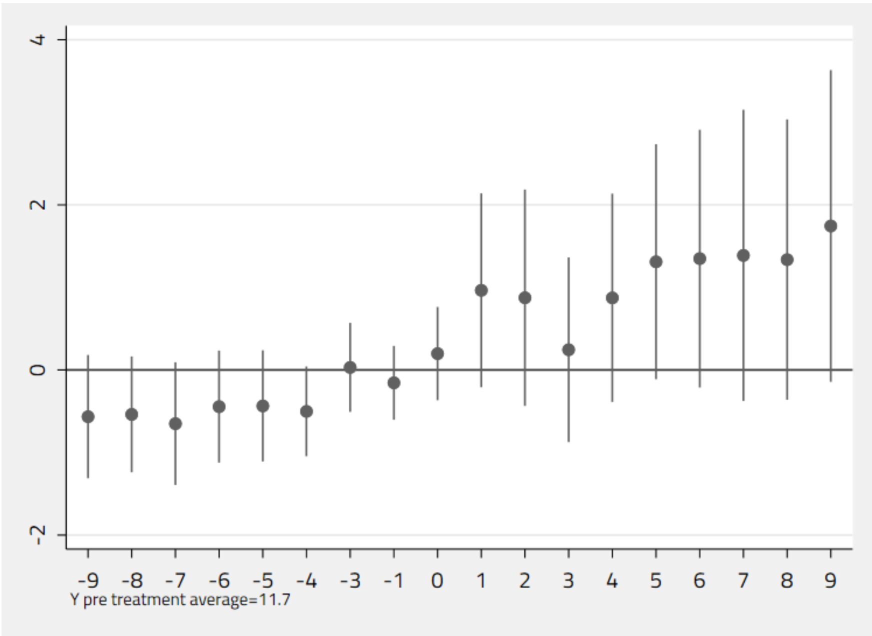


Figure 10: Workers Employed around the Abolition of Vouchers

Notes: The figure plots differences in the total number of workers employed at firms that on average "mis-used" 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.

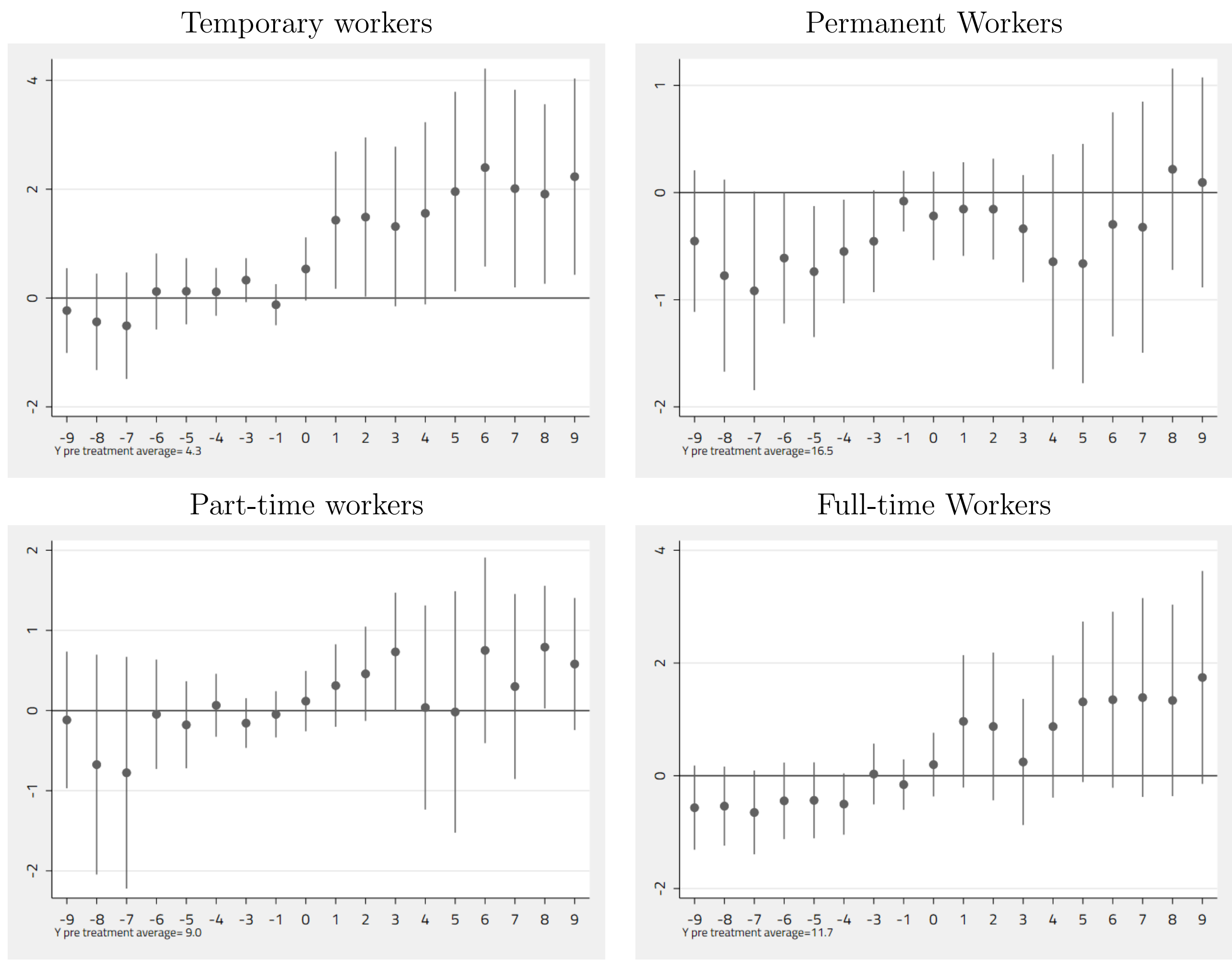


Figure 11: Event Study around the Abolition of Vouchers

Notes: The figure plots differences in the number of workers employed at firms that on average "mis-used" 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.

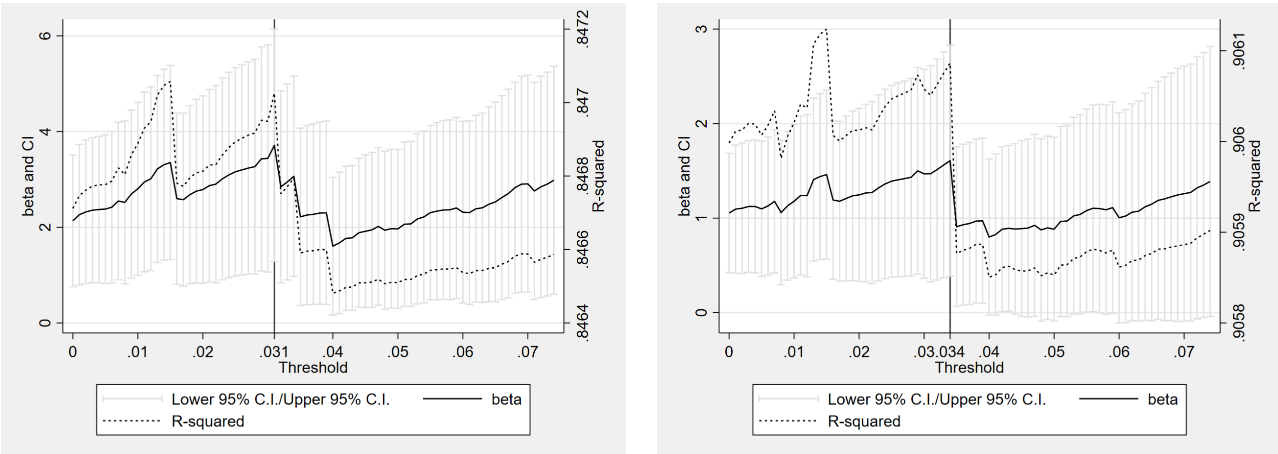


Figure 12: Sensitivity analysis on the misclassification bias

Notes: The figures plot differences in the total number of temporary workers (left panel) as well as temporary part-time workers (right panel) employed at firms that on average "mis-used" vouchers and those that did not, and the corresponding 95% confidence intervals. The definition of misusing firm $\widehat{M}_j^\eta$ varies based on the η threshold: $\widehat{M}_j^\eta = 1\{\hat{\beta}_j > \eta\}$. Standard errors are clustered at the firm level.

Table 1: Summary Statistics

Sample	Full	Vouchers	Inspected	Both	Jump up	Jump down
Workforce	7.18	7.27	43.49	40.41	36.55	45.98
Monthly Wage bill	14,387	12,131	95,185	62,261	52,958	75,704
Average wage	1,347	1,273	1,316	1,248	1,249	1,246
Monthly Wage bill (FTE)	15,935	13,891	103,619	74,550	65,846	87,128
Workforce (FTE)	6.27	6.21	38.75	33.36	29.21	39.35
Temporary workers	1.10	1.81	8.51	16.44	11.70	23.29
Part-time workers	2.13	2.68	11.13	16.90	17.15	16.53
Permanent workers	6.08	5.46	34.97	23.97	24.85	22.69
Full-time workers	5.05	4.59	32.36	23.51	19.40	29.45
Monthly number of vouchers	27.75	97.60	299.78	251.91	269.53	238.97
Voucher per workforce	4.43	15.73	7.74	7.55	9.23	6.07
Number of firms	1,836,191	416,930	29,063	3,472	1,471	2,001
Fraction	100%	22.7%	1.58%	0.19%	0.08%	0.11%

Notes: This table shows averages for different samples using yearly 2015 and 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 (almost all are inspected only once); “Both” are firms that use at least on voucher and have been inspected. “Jump up” and “Jump down” are firms that either increase of decrease the use of vouchers upon inspection. “FTE” stands for full-time-equivalent

Table 2: Vouchers and labor 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.

Table 3: Vouchers and labor Inspections Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Subsample	South	Center	Nort-East	North-West	Young firms	Medium-aged f.	Old firms	Small firms	Medium-sized f.	Large firms	Above-median use of PT
Post-Inspection	0.009**	0.008**	0.010***	0.011***	0.009***	0.011***	0.008**	0.006***	0.008***	0.012***	0.014***
	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.003)	(0.004)	(0.003)
Constant	0.047***	0.046***	0.047***	0.043***	0.045***	0.048***	0.044***	0.035***	0.044***	0.058***	0.057***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.001)	(0.002)	(0.003)	(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
R-squared	0.015	0.015	0.014	0.014	0.014	0.013	0.017	0.012	0.016	0.017	0.005
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

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: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. 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.

Table 4: Vouchers and Evaded Contributions

	(1)	(2)	(3)	(4)
Post-Abolition*Treated	27,979*	30,423**	34,626**	32,968*
	(15,351)	(15,265)	(16,874)	(16,868)
Post-Sms*Treated	-15,174**	-15,402**	-12,536	-14,354*
	(7,125)	(7,121)	(8,005)	(8,037)
Pre-Sms*Treated	-8,421	-8,589	-4,940	-5,416
	(6,573)	(6,519)	(6,681)	(6,685)
Treated	3,889	4,049	394	1,183
	(4,430)	(4,431)	(4,709)	(4,723)
Post-Abolition	23,673***	28,974***	24,610***	24,513***
	(6,095)	(7,492)	(8,022)	(8,218)
Post-Sms	17,505***	11,943**	6,375	7,155
	(5,240)	(5,660)	(6,088)	(5,981)
Pre-Sms	15,516***	22,758***	16,314**	16,476**
	(4,662)	(7,327)	(6,479)	(6,455)
Observations	20,819	20,819	18,109	18,109
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. The baseline period is January to October 2016. Treated firms have used at least one voucher during the sample period, control firms have never used vouchers. The Pre-SMS period goes from October 2016 to March 2017, while the Post-Abolition period starts in March 2017. Standard errors are clustered at the firm level. *** p<0.01, ** p<0.05, * p<0.1.

A Appendix

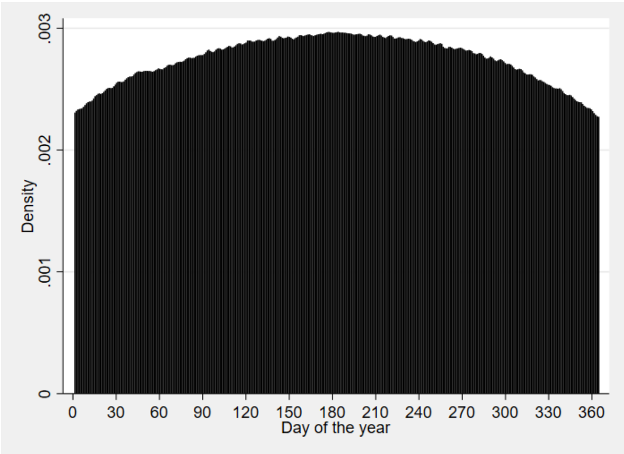
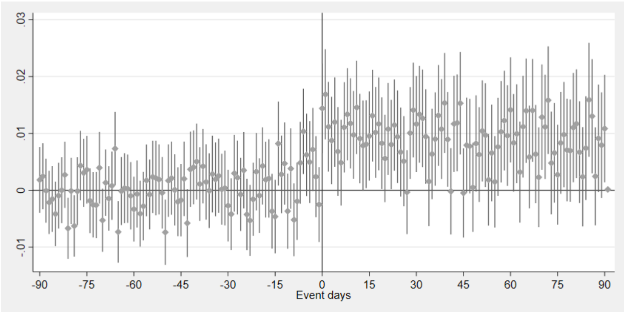


Figure A1: Distribution of Labor Inspections

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



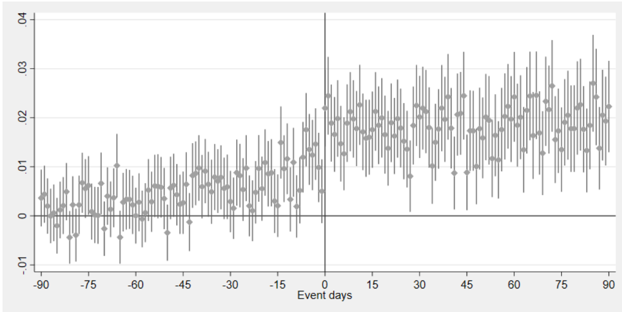


Figure A2: Event Study pre-SMS

Notes: The figure plots event study coefficients, where the event is a labor 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. 11, $f(t) = \alpha t$), the figure on the right has no additional controls (in Eq. 11, $f(t) = 0$).

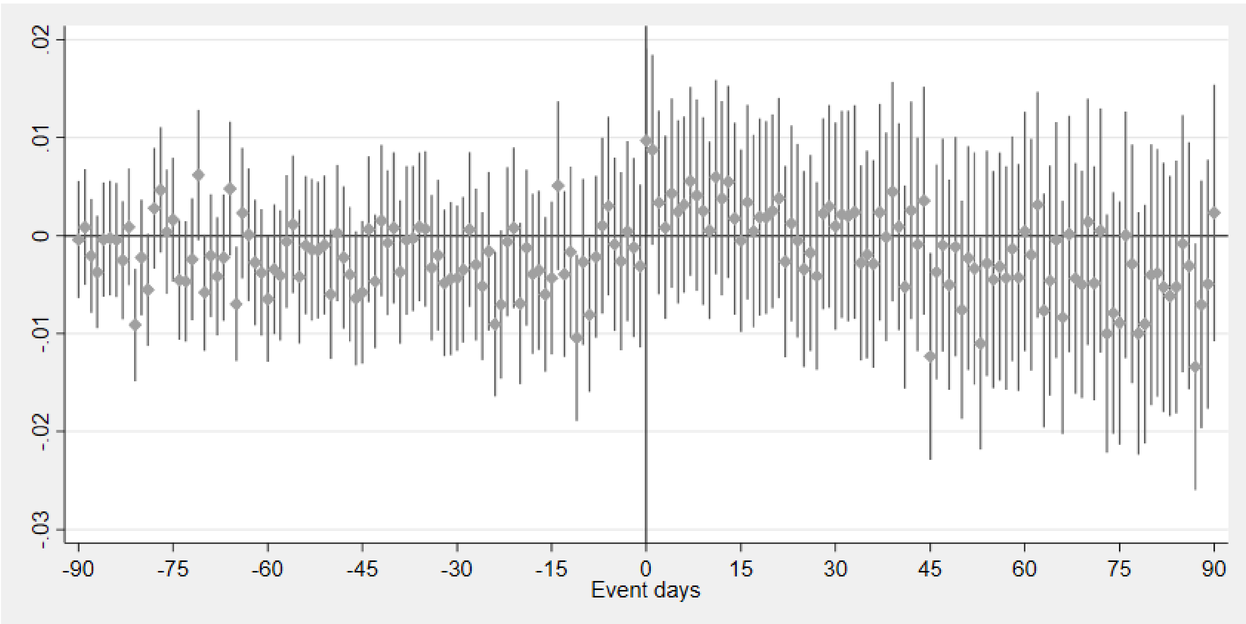


Figure A3: Event Study pre-SMS with Firm Fixed Effects

Notes: The figure plots event study coefficients, where the event is a labor 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.

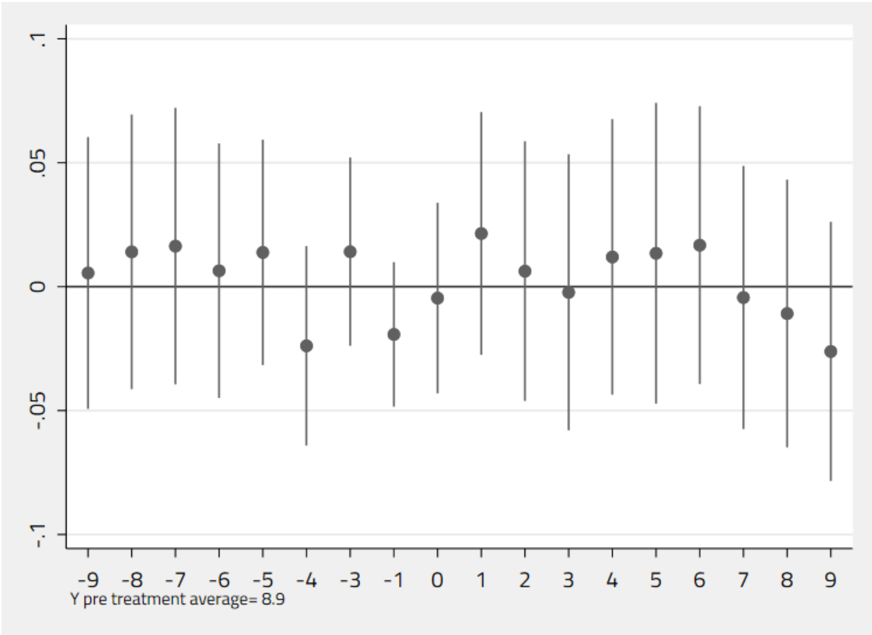


Figure A4: Log-Total Wage Bill Around the Abolition of Vouchers

Notes: The figure plots differences in the total wage bill between firms that on average "mis-used" 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: Difference-in-Differences Models of Vouchers and labor 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.

Figure A5: The Photograph of an Italian voucher