



DEPARTMENT OF ECONOMICS,  
MANAGEMENT AND STATISTICS  
UNIVERSITY OF MILAN – BICOCCA

## DEMS WORKING PAPER SERIES

### **Is Tax Compliance a Social Norm? A Field Experiment**

Pietro Battiston, Simona Gamba  
No. 249 – July 2013

Dipartimento di Economia, Metodi Quantitativi e Strategie di Impresa  
Università degli Studi di Milano - Bicocca  
<http://dems.unimib.it/>

# Is Tax Compliance a Social Norm? A Field Experiment\*

Pietro Battiston<sup>†</sup> and Simona Gamba<sup>‡</sup>

June 2013

## Abstract

We study the effect of social pressure on tax compliance, focusing on the compliance of shop sellers to the legal obligation of releasing tax receipts for each sale. We carry out a field experiment on bakeries in Italy, where a strong gap exists between the legal obligation and the actual behavior of sellers. Social pressure is manipulated by means of an explicit request for a receipt when not released. We find that a single request for a receipt causes a 17 per cent rise in the probability of a receipt being released for a sale occurring shortly thereafter. This provides evidence of a social fiscal multiplier: on average, a single request for a receipt causes 2.38 additional receipts being released overall.

**Keywords:** Tax evasion, field experiments, social norms, social pressure.

**JEL classification:** H32, K34, E62

---

\*We thank Luca Stanca and Alessandro Santoro for providing fruitful suggestions and continuous support. We thank Claudio Lucifora for helping us to clarify our empirical strategy. We are grateful to Björn Frank, and to all others participants at conferences where this work was presented, for their insightful comments.

<sup>†</sup>DEFAP, Department of Economics, University of Milano Bicocca, Piazza Ateneo Nuovo, 1, 20126 Milano, Italy. me@pietrobattiston.it

<sup>‡</sup>DEFAP, Department of Economics, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123 Milano, Italy. gamba.simona@gmail.com

# 1 Introduction

The literature on fiscal compliance has developed from the seminal model of Allingham and Sandmo (1972). In their work (as in similar studies by Kolm, 1973, and Singh, 1973), expected utility maximizing agents choose the income level to be reported to the fiscal authority, considering the probability of being audited and the size of the fine. At the empirical level, however, researchers have faced a major puzzle: in all advanced economies, the level of tax compliance is far higher than the one predicted by the theory (e.g. Graetz and Wilde, 1985, Alm et al., 1992).

A stream of literature has approached this discrepancy by extending the original model with more realistic specifications of the context in which tax declaration choices are made. In this context, financial strain (Wärneryd and Walerud, 1982) and the broad category of *opportunities* have been analyzed, among other factors. The role of third-party reporting, which limits the possibility of employees to evade taxes, has been widely discussed (Andreoni et al., 1998) and tested experimentally (e.g. Slemrod, 2007 and Kleven et al., 2011).

Even these studies, however, recognize that the high level of compliance empirically observed cannot be fully explained without taking into account behavioral factors. This observation has motivated further extensions of the basic model. Bordignon (1993) embeds fairness-based evaluations into the utility function, while Gordon (1989) introduces non-pecuniary *stigma* costs associated with tax evasion. Weigel et al. (1987) and Groenland and Van Veldhoven (1983) provide a *social and psychological model*, which represents a broader approach to the several conditions which influence fiscal behavior, such as personality (see Lewis, 2011). Studies on behavioral aspects of tax compliance are rooted in the wider stream of literature about the social aspects of deterrence (see for instance Grasmick and Bursik Jr, 1990 and Paternoster et al., 1983).

One particular aspect that has been of interest for a long time is the definition and relevance of the *reference group* of an individual: “relevant others” are the members of a group where the principle of reciprocity applies (Kirchler, 2007 pp. 64-65). Two recent studies that feature this aspect in a prominent role are Wenzel (2004) and Galbiati and Zanella (2012). Erard and Feinstein (1994) merge the complementary approaches of utility maximization and tax morale by exploring the consequences of an (exogenously given) share of *honest* taxpayers on the audit rate for other citizens - and

hence on their fiscal behavior. In equilibrium, even purely selfish citizens end up paying more than in the basic framework.

Studies of tax compliance are generally confronted with a lack of data that is particularly hard to overcome, as effectively summarized in Cowell (1991): “Data from official investigations are hardly ever available and data from other sources may be suspect: if you could directly observe and measure a hidden activity, then presumably it could not really have been properly hidden in the first place.”

Weigel et al. (1987) considered as fundamental for future fiscal research the development of *creative methods* for attaining objective estimates of tax evasion behavior. The issue is still open, as reported in recent years by Halla (2012). In particular, the frequent use of survey data, where individuals self-report their tax behavior, has since long been perceived as a crucial issue (Weigel et al., 1987, Elffers et al., 1987), because of the possible misreporting.

Therefore, a growing stream of literature has focused on *experiments* aimed at reproducing the economic and psychological reasoning behind tax compliance. This stream can be traced back to Reis and Gruzen (1976) and Kidder et al. (1977). More recent attempts in this direction are those of Alm et al. (1992) and Cummings et al. (2006). In his exhaustive review of the field, Torgler (2002) acknowledges the relevance of experiments in that tax enforcement, tax rate and income levels can be controlled.

The effect of social norms and social disapproval on the tax compliance choice, in particular, has been approached experimentally in Cummings et al. (2001), Alm et al. (2007), and Bosco and Mittone (1997). The latter tests two separate hypotheses, concerning the effects of either *subjective* or *collective* moral constraints on tax compliance. Subjective moral constraints are manipulated as follows: while in the control group money collected through taxes is just taken away from participants, in the experimental treatment there is a partial redistribution of the collected amount. In order to test for the second hypothesis, instead, a treatment is run in which the identity of individuals who are caught cheating is publicly revealed, and evaders hence run the risk of being identified as such by other participants. The authors find significant evidence only in favor of the first hypothesis.

However, Torgler (2002) also casts doubts on the fact that laboratory experiments can be considered informative about actual tax compliance behavior. This view is shared by Halla (2012), who suggests that individuals react to experimenters’ stimuli differently than with real tax authorities. Social norms and the related aspect of social pressure, are probably among the

determinants of tax compliance which are *least reproducible* in a laboratory. Nevertheless, they are a fundamental ingredient of the compliance decision (Posner, 2000) because they “constitute constraints on individual behavior beyond the legal, information and budget constraints usually considered by economists” (Fehr et al., 2002).<sup>1</sup>

Relatively few attempts have been made to identify the size and the determinants of tax evasion through the use of field experiments. This is due in part to the typical reluctance of national fiscal authorities toward randomized actions, which are supposed to go against the principle of equity.<sup>2</sup> The approach of Schwartz and Orleans (1967), later adopted by Wenzel (2001), is based on surveys sent to taxpayers some time before they file their tax declaration. The questions asked vary from group to group. This enables the authors to identify the reduction in evasion due to “*conscience*” versus the one due to “*sanctions*”, by arousing respectively the feeling of *guilt* related to the social loss or the *fear* of detection. While they find that the relative importance of the two motives depends on the social and economic status of individuals, overall they report that “*conscience appeals are more effective than sanction threats*”. Slemrod et al. (2001), through threat-of-audit letters, identify the response of taxpayers to an increase in audit probability, and report mixed evidence. They find an increase in amounts declared by low and middle-income taxpayers, but a *decrease* in amounts declared by high-income ones. This result is attributed to the particular wording used in the letters, together with the heterogeneity of beliefs and of information that individuals have about the fiscal authority. Kleven et al. (2011) bring into the picture the effects of an audit itself on subsequent tax declarations, as an indicator of undeclared income. Their main conclusion is that fiscal evasion is severely hindered by third-party reporting. Still, they acknowledge the evidence of behavioral factors: even though audits do not imply a higher audit probability in the future, they have a positive deterrence effect on the following fiscal year.

All these studies focus on the *institutional* determinants of the compliance decision, by randomizing the sample of taxpayers to whom surveys, audits, or

---

<sup>1</sup>Fehr et al. (2002) define social norms as “(i) a behavioral regularity that is (ii) based on a socially shared belief regarding how one ought to behave which triggers (iii) the enforcement of the prescribed behavior by informal social sanctions.”

<sup>2</sup>Randomized setups are characterized precisely by the fact that they treat equal citizens differently, rather than shaping enforcement actions deterministically on observable variables.

threat-of-audit letters, are assigned. Instead, to the best of our knowledge, no field experiments have been implemented with a focus on the effect of social pressure. The present paper tries to fill the gap between theory and evidence. It does so by exploiting the particular case of tax evasion among shop sellers in Italy, a country where non-compliance is relatively widespread (concerning at least 30% of bakeries in our estimates). The vast majority of Italian shops are obliged by the law to release a tax receipt for each sale.<sup>3</sup> The total sum of receipt amounts represents the revenues of a shop, and receipts themselves can constitute a proof for the fiscal authority. Both value added tax and income tax are calculated on the basis of such revenues. As a consequence, the omitted release of a receipt is an act of fiscal evasion, allowing the seller to evade both value added and income tax. Interestingly, this act of tax evasion is not only common, but also, at least in the case under analysis, committed *openly*, making it trivial for a purchaser to ascertain non-compliance. Although it would also be trivial for the purchaser to actively *fight* tax evasion - by simply requesting the receipt when it is not released - this behavior is far from being widespread. When such a request is made, it can be seen as an implementation of a social norm.

According to Kirchler (2007) social norms are a function of an individual's perceived expectation that one or more relevant referents would approve a particular behavior. In the prevailing literature, such "relevant others" are members of the group where the principle of reciprocity applies; the concept of *strong reciprocity*<sup>4</sup> and its importance for the enforcement of social norms are studied in detail by Fehr and Gächter (1998) and Fehr et al. (2002). In this empirical study, the reference group can be seen as composed by occasional customers: they are, for the vendor, representatives of the "common view" of the Italian society. Thus, a customer who, after a purchase, accepts to leave the shop without tax receipt is a relevant other who is tolerating the tax evading behavior. Conversely, the customer who requests the tax receipt is blaming the vendor for his behavior and is enforcing the social rule of tax compliance.

The particular activity we study exhibits some features that make it suitable for the experiment. First, the good at sale, bread, is relatively stan-

---

<sup>3</sup>Supermarkets and kiosks are among the few exceptions, which however are irrelevant in this context.

<sup>4</sup>"A person is a strong reciprocator if she is willing to sacrifice resources (a) to be kind to those who are being kind (strong positive reciprocity) and (b) to punish those who are being unkind (strong negative reciprocity)."

standardized, making it easier to compare different shops. Second, it has a low cost, which implies that the profit obtained by evading is generally not the object of bargaining between the seller and the buyer.<sup>5</sup> Therefore, the act of requesting the receipt does not directly affect the utility of the buyer: it can instead be considered a pure act of social pressure.<sup>6</sup>

The remainder of the paper is structured as follows. Section 2 presents the theoretical basis for our analysis. We then describe our experimental setting (Section 3). Section 4 provides details on our identification strategy and Section 5 presents the results. Finally, Section 6 summarizes our conclusions.

## 2 Theoretical background

Bordignon (1993) bases its explanation of “excessive” tax compliance on *fairness-based* motives: citizens do not only care about their own net income, but also about social welfare. While this approach certainly brings into the picture an interesting aspect of taxpayers’ decision making, its possible implications in terms of social pressure are non obvious.<sup>7</sup>

In the model of Gordon (1989), instead, the utility of the taxpayer takes the general form

$$u(C, H)$$

where  $C$  is the ordinary consumption, while  $H = -E$ , and  $E$  is the amount of undeclared income (embodying the concept of *stigma*). Clearly, consumption depends on the amount of income concealed:

$$C = Y(1 - \tau) + x\tau E$$

with  $Y$  representing disposable income, and  $\tau$  the tax rate.  $x = -s$  if the cheater is caught,  $x = 1$  otherwise. Since  $u$  is increasing in both arguments,

---

<sup>5</sup>Bargaining is far from being unrealistic in other sectors, in which the amount of evasion benefits *per purchase* is much higher and the buyer is often offered a discount conditional on not receiving the receipt.

<sup>6</sup>The Italian legislation used to punish clients unable to show the receipt of a purchase just made. This law, however, was abolished in 2003, leaving only the existing sanctions on sellers.

<sup>7</sup>Notice that in principle the effect of social pressure could partly overlap with fairness motives. It could be interpreted by the potential evader as a signal of private information on the trustworthiness of the public government - and hence of the good use of public revenues. This could be a driving factor toward higher compliance for a fairness-motivated seller.

this clearly implies a trade-off between the *stigma cost* and the cost of compliance: under given functional assumptions, the utility function is rewritten as

$$u(C, E) = U(C) - vE$$

where  $vE$  is the *private psychic cost of evasion*.  $v$  is assumed to be distributed in the population according to some distribution  $F(v)$ .

The choice of relying on the concept of *stigma* has the interesting implication that individuals may pay more taxes than what would be optimal for their balance sheet, *in particular* if they are more afraid to be caught (even keeping fixed the expected fine).

We extend the model by enriching the utility function. In particular, we assume that the psychic cost for the taxpayer depends on the *ethical tastes* of the peers:

$$\tilde{u}(C, E) = U(C) - f(\pi)E$$

where  $\pi$  captures the extent to which the individual is concerned with the social attitude against tax non-compliance, and  $f$  is a strictly increasing function. Without loss of generality, we set  $f(0) = v$ .<sup>8</sup>

The optimal level of tax evasion  $e^*$  when  $\pi = 0$  is, by definition, the same as in the original model by Gordon. Let us consider now the case where  $\pi > 0$ . We can write  $\tilde{u} = u - (f(\pi) - v)E$ , and hence

$$\frac{\partial \tilde{u}}{\partial E}(e^*) = \frac{\partial u}{\partial E}(e^*) - (f(\pi) - v).$$

By definition of  $e^*$ , we have that  $\frac{\partial u}{\partial E}(e^*) = 0$ , and hence

$$\frac{\partial \tilde{u}}{\partial E}(e^*) = -(f(\pi) - v) = -(f(\pi) - f(0)) < 0;$$

that means  $e^*$  now represents a sub-optimal level of evasion: the new optimal level will be  $\tilde{e}^* < e^*$ . It is easy to verify that the starting condition  $\pi > 0$  is both sufficient and necessary for this conclusion, and that in general  $\tilde{e}^*$  decreases for increasing values  $\pi$ .

---

<sup>8</sup>For  $\pi$  to be 0, it is sufficient that the society does not take a stand against tax evasion, or that it does, but the seller is unconcerned or unaware. In this latter case, a request for the receipt could represent an information on previously *unknown* ethical tastes, or updated information about a recent *change* in such tastes.



The aim of our experiment is to verify empirically this prediction, by studying the variation in  $\tilde{c}^*$  when manipulating  $\pi$ . Moreover, the experiment may shed some light on an additional hypothesis suggested by the model: “individuals cheat less if they are observed by *more* peers”. Notice that the model does not give an unambiguous prediction concerning this hypothesis: the effect of peers simply standing by while an individual makes her compliance choice is unclear: it depends on her belief concerning the “ethical tastes” of those peers.

### 3 The experiment

The experimental sample consisted of 108 bakeries located in the city of Milan. For each bakery, the time line of the experiment was articulated in two periods. In period 1, an agent entered the shop and bought a loaf of bread. The purchase was paid with an amount of money higher than its cost,<sup>9</sup> so that the agent had to wait for the change. This design choice was made because a client standing in front of the seller after having paid and having received the bread could have influenced his behavior. In this way instead, the moment in which the change was given (with or without the receipt) represented unambiguously the end of the transaction. Only at this time, if the receipt had not been released, the agent asked for it:<sup>10</sup> this request was our treatment.<sup>11</sup>

In period 2, twelve minutes after the first agent left the shop, another agent entered the same bakery. Following the same procedure as in period 1, the agent bought a loaf of bread of a different type.<sup>12</sup> The role of this second agent was to assess if the receipt was now given. Whatever was the behavior of the seller, no request for a receipt took place at this time.

The choice of the 12 minute time span was made because it is absolutely unlikely that any client would spend such an amount of time in a bakery.

---

<sup>9</sup>For the sake of homogeneity, we also never paid with banknotes, and the amount given was always lower than 2 €.

<sup>10</sup>The receipt was always requested using the same wording (“*Vorrebbe essere così gentile da rilasciarmi lo scontrino?*”), which roughly translates to “*Would you be so kind as to give me the receipt?*”).

<sup>11</sup>Note that the request for a receipt was made only for non-compliant bakeries.

<sup>12</sup>The change of agent and of type of bread being asked, together with the fact that bakeries are characterized by a high number of low volume sales, makes the whole interaction perfectly natural.

This means that when the second agent entered, the first one, as well as the clients present in the shop when the request for the receipt had taken place, had already left the shop. In this way, any change in the behavior of sellers can be attributed uniquely to a reaction of the sellers themselves to the request, as opposed to *indirect* pressure, or to the presence of the client who proved to be particularly “picky”.

All non-compliant bakeries were treated, in order to extract as much information as possible from the intrinsically small sub-sample consisting of bakeries *of a given city* which *at a given time* abstain from releasing the receipt. In Section 4 we show that this non-randomness of the treatment group does not restrict our ability to attribute a causal interpretation to our estimates.

Since the experiment consisted of buying twice from each of the bakeries and then comparing the results in terms of compliance, we made sure that each of the two passes had, on average, exactly the same characteristics (except for the treatment), in order to remove any potential confounding factor. Therefore, (a) the entry order of the two agents (one male and one female, both around 25 years of age) was randomized; (b) the types of bread purchased were randomized;<sup>13</sup> most importantly (c), the second agent *did not know* if the first had been spontaneously given the receipt and hence if the bakery had been treated. Before proceeding to the analysis of our data, it should be pointed out that the two agents could possibly meet two different sellers inside the shop. We found that this is an unlikely event: in more than 60% of the shops, only one seller was present during both transactions.<sup>14</sup> Still, there were shops (at least 14% of them, as evidenced from the non-matching sellers characteristics) in which the vendor did change between the two passes, making our estimates *within bakery* rather than *within seller*. These estimates are what matters for policy implications (social pressure is still relevant even if felt indirectly by another seller of the shop) and presumably represent a lower bound for the *within seller* effect.

In order to obtain a preliminary assessment of tax compliance, a single

---

<sup>13</sup>Each time, one agent asked for a type of bread and the other one asked for another, resorting to a third and then to other types if the requested one was not available. The three types chosen are comparable in weight, size, cost, and all of them are usually sold by any bakery.

<sup>14</sup>We also verify that when this was the case, the characteristics (gender and apparent age) of the vendor recorded in the first and in the second pass were the same in 97% of cases.

pass was carried out in January 2012, with 177 bakeries being investigated. Of these, 22% did not release a receipt to the agents.<sup>15</sup> Given the purely descriptive purpose of this first investigation, no treatment was administered. The field experiment took place shortly after, in the first 2 weeks of March 2012. At that time, all the bakeries that had not released the receipt during the preliminary investigation were inspected, together with 70 other bakeries randomly selected. In total, 108 shops were included in the experiment: 21% did not release the receipt during the first pass. Of these, 13 were treated by a female agent, 10 by a male agent - each agent had entered as first in exactly 50% of the total bakeries. Among the treated bakeries, one type of bread had been asked in 11 cases, and the other type in 12 cases. A summary of the compliance outcomes which will be relevant for the analysis is presented in Table 2.

## 4 Methods

We are interested in estimating the causal effect of social pressure on tax compliance. The population of interest for our experiment is the set of all bakeries in the city of Milan. The treatment (denoted by the Boolean variable  $D_i$ ) consists of requesting the receipt if not released spontaneously: therefore, only non-compliant bakeries are treated. In the terminology of the classic treatment-effect framework, the aim of our analysis is the estimation of the Average Treatment on the Treated (ATT), answering the following question: “does exerting social pressure *on non-abiding sellers* affect their propensity to tax compliance?” In fact, it is hardly possible to measure an ATE (Average Treatment Effect), since it is difficult to imagine treating bakeries where the receipt is spontaneously given.<sup>16</sup>

---

<sup>15</sup> Additionally, at least 15 other bakeries (8.5%) released a receipt to the agents but *not* to some other client present at the same time, for a total of at least 30.5% of bakeries observed in a non law-abiding behavior.

<sup>16</sup>In principle, an experiment could be ran in which social pressure is exerted *at the start* of the transaction, for instance with agents stating explicitly they want the receipt at the moment of asking the loaf of bread. This experimental setup, by randomizing the sample of treated bakeries, would indeed produce an ATE. This was not our choice for two reasons. First, the measurable effects would have been largely diluted. Second, the practical implementation of our experiment is much more realistic: it does happen, although it is certainly not the norm, that clients not receiving the receipt ask for it, while it is much less common that a client asks for the receipt beforehand.

Table 1: Relevant probabilities

		Second pass ( $c^2$ )		
		0	1	Total
First pass( $c^1$ )	0	$\mathbb{P}_D\{c^1 = 0, c^2 = 0\}$	$\mathbb{P}_D\{c^1 = 0, c^2 = 1\}$	$\mathbb{P}\{c^1 = 0\}$
	1	$\mathbb{P}\{c^1 = 1, c^2 = 0\}$	$\mathbb{P}\{c^1 = 1, c^2 = 1\}$	$\mathbb{P}\{c^1 = 1\}$

Table 1 introduces the moments of data on which our analysis is based. The compliance status observed in a generic bakery during pass  $j$  ( $j = 1, 2$ ) is denoted as  $c_i^j$ . This is a Boolean variable, equal to 1 if and only if the bakery did release the receipt.  $\mathbb{P}_D$  denotes probabilities for treated bakeries, while  $\mathbb{P}$  represents the “natural” probability, in absence of any treatment.

The identification of the ATT is equivalent to finding a causal relationship between  $D_i$  and  $c_i^2$ , restricting the attention to bakeries with  $c_i^1 = 0$ . In principle, we could obtain this by estimating a simple regression of the outcome ( $c_i^2$ ) on the treatment status:

$$c_i^2 = \beta_0 + \beta_1 D_i + u_i. \quad (1)$$

Notice, however, that having treated all non-compliant bakeries,  $D$  is perfectly collinear with  $c^1$ :

$$D_i \equiv 1 - c_i^1. \quad (2)$$

Hence, equation 1 cannot be estimated in a straightforward way in the set  $\{c_i^1 = 0\}$ . In what follows, we describe our identification strategy. By definition, the Average Treatment Effect on the Treated can be expressed as:

$$ATT = \mathbb{E}\{c^2 | c^1 = 0, D = 1\} - \mathbb{E}\{c^2 | c^1 = 0, D = 0\}. \quad (3)$$

Notice that the only difference between the two terms on the right hand side is the treatment status, so that we can rewrite the above as:

$$\frac{\mathbb{P}_D\{c^1 = 0, c^2 = 1\}}{\mathbb{P}\{c^1 = 0\}} - \frac{\mathbb{P}\{c^1 = 0, c^2 = 1\}}{\mathbb{P}\{c^1 = 0\}} = \frac{\mathbb{P}_D\{c^1 = 0, c^2 = 1\} - \mathbb{P}\{c^1 = 0, c^2 = 1\}}{\mathbb{P}\{c^1 = 0\}}.$$

We mentioned in Section 3 that the two passes were designed in order to be perfectly comparable. Intuitively, this means that the frequency of the event “compliance only at the first sale” ( $c_1 = 1, c_2 = 0$ ) is expected to be

the same as the frequency of “compliance only at the second sale” ( $c_1 = 0$ ,  $c_2 = 1$ ):

$$\mathbb{P}\{c^1 = 0, c^2 = 1\} = \mathbb{P}\{c^1 = 1, c^2 = 0\}. \quad (4)$$

A detailed justification of this equation can be found in Appendix B.

We can estimate the second term of equation (4) from data for bakeries that released the receipt at the first pass ( $c^1 = 1$ ), and hence were not treated. As a consequence, equation (3) can be rewritten as

$$ATT = \frac{\mathbb{P}_D\{c^1 = 0, c^2 = 1\} - \mathbb{P}\{c^1 = 1, c^2 = 0\}}{\mathbb{P}\{c^1 = 0\}}. \quad (5)$$

All terms appearing in (5) can be estimated from our data.

## 5 Results

Table 2 summarizes the experimental data, providing the empirical realizations corresponding to cases in Table 1. Notice that the frequency compliance during the second pass ( $c^2 = 0$ ) was 81%, higher than during the first pass.

Table 2: Summary of data

		Second pass ( $c^2$ )		
		0	1	Total
First pass ( $c^1$ )	0	7	16	23 (21 %)
	1	12	73	85 (79%)
Total		19 (18%)	89 (82%)	108 (100%)

By estimating (5) we obtain:

$$ATT \approx \frac{\frac{16}{108} - \frac{12}{108}}{\frac{23}{108}} = 17.4\%.$$

The treatment lowers the probability of not receiving the receipt in the second pass by 17.4 points: it decreases from 47.8% to 30.4%. Notice that both values are higher than the non-compliance rate we expect from a generic bakery, which is 21% (see Table 2, and recall that for an *untreated* bakery the two passes are perfectly equivalent). Hence, our data show that bakeries which are non-compliant in the first pass are *intrinsically* different from the

others, thus providing strong evidence of illegal behavior persistence. This phenomenon is so strong that non-compliance in the first pass increases the probability of non-compliance in the second *even when* the request for a receipt takes place.<sup>17</sup>

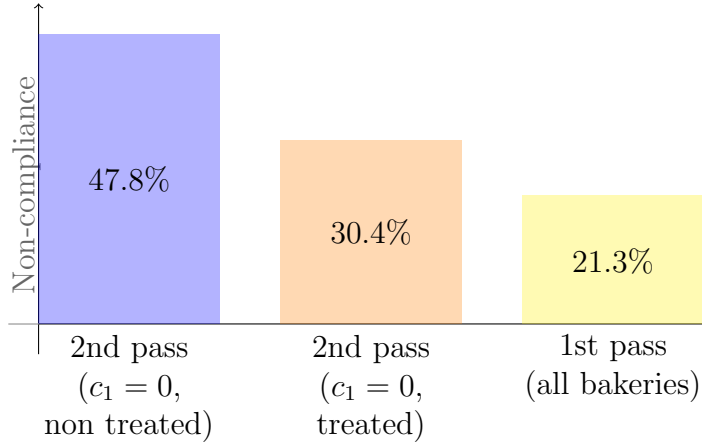


Figure 1: Predicted persistence of non-compliance without (blue) and with treatment (orange), compared to the probability of non-compliance of a generic bakery in the first round (yellow).

Interestingly, the effect of the treatment seems to be stronger among bakeries which were the object of the preliminary investigation discussed at the beginning of Section 3 (and which were at the time non-compliant): in such a sub-sample, the estimation of the ATT is 36.36%. This seems to suggest that the request for the receipt has a stronger impact on bakeries which are *frequently* non-compliant.

In order to assess if the estimated ATT is statistically significant, we run a one-sided test of  $\mathcal{H}_1 : ATT > 0$  against the null hypothesis  $\mathcal{H}_0 : ATT = 0$ . Since the small sample makes asymptotic distributional assumptions unlikely, we refrain from using McNemar’s test to reject the possibility of a type I error.<sup>18</sup> As suggested in Sheskin (2004), we instead observe that, assuming  $\mathcal{H}_0$  holds,  $N_{01}$  is distributed according to a binomial:

<sup>17</sup>Running an exact Fisher test for Table 2 allows us to reject the null hypothesis that the first and the second pass are independent ( $p = 0.003$ ), hence proving that fiscal behavior is characterized by persistence, *despite* our treatment effect.

<sup>18</sup>We also refrain from using a Fisher exact test, since we do not have a complete contingency table (we do not observe  $\mathbb{P}\{c^2 = 0 | D = 0\}$ ).

$$N_{01} \sim \mathcal{B}(p, N_0)$$

where  $p$  is the probability of changing compliance status by pure chance. Denoting as  $\hat{p}$  our estimate of  $p$ , the probability of a type I error is therefore:

$$\mathbb{P}\{N_{01} \geq 16 | N_{01} \sim \mathcal{B}(\hat{p}, N_0)\} = 7\%$$

(see Appendix A for details). We are hence able to identify a causal effect of the treatment on the treated (*ATT*) at a significance level of  $\alpha = 10\%$ .

The same procedure restricted to the sub-sample of bakeries which had already been visited in the January 2012 preliminary investigation (and had not, at the time, released the receipt) gives a  $p$ -value of 0.015. When restricting to bakeries *not* visited in the preliminary pass, instead, the results are not significant. This is likely to reflect the small sample size.

## 5.1 A social fiscal multiplier

Consider a client not receiving a receipt and asking for it. The *ATT* measures the effect of this event on the probability that, approximately 12 minutes later, another client receives a receipt. It is interesting to find out *how many* receipts can be expected to be released, overall, as a consequence of that single request.

On average, treated bakeries released 5.5 receipts between the two passes.<sup>19</sup> This is clearly a lower bound for the number of clients they had in this time span. In fact, let us assume that the attitude towards releasing receipts is homogeneous across (rebuked) bakeries and across clients during the first 12 minutes. Then, the observed non-compliance rate (Figure 1) calls for a number of clients  $\eta$  such that:

$$\eta \cdot (1 - 30.4\%) = 5.5 \implies \eta = 7.9.$$

Now, if  $\eta$  clients enter the bakery in the 12 minutes and the *ATT* at the end of this period is 17.4%, then we expect that at least

$$\eta \cdot ATT \approx 1.38$$

---

<sup>19</sup>This figure can be obtained from the subset of bakeries having released the receipt in the second pass. This exercise hence implicitly assumes negligible correlation between the number of clients and the persistence of non-compliance.

additional receipts are released because of the treatment. This suggests the existence of a *social fiscal multiplier*: each rebuke causes the release of 2.38 receipts in total. This estimate can be viewed as a lower bound for two reasons. First, sales occurring *before* the 12th minute are expected to be affected *even more* by the rebuke (assuming its effect fades away with time), and this produces a downward bias in our estimate of additional receipts;<sup>20</sup> second, the calculation above ignores the possible effects on sales occurring *after* the 12th minute.

However, since we are looking for a *general* effect, some further issues must be kept in mind. The effect of the treatment may be *local*: for instance, a seller who has been rebuked by a young client may, in the future, increase compliance when facing young clients. Moreover, the agents had no ambitions of representing the average client of a bakery in terms of observable characteristics and loyalty to the shop. While it is reasonable to think that a rebuke coming from a loyal customer will presumably have an *even higher* psychological impact on the seller, we have no way of predicting the effect of other variables, such as age. This could be an interesting topic for further research. Finally, the sample under observation is not purely random, in the sense that 38 of the 108 bakeries were chosen for their non-compliant behavior during the January 2012 preliminary investigation. Further research could be devoted to the interplay between persistence of illegal behavior and reaction to social pressure.

## 5.2 Robustness

When the client and the vendor are of the same gender, the probability of the receipt being released drops by 13.6% and the difference is statistically significant at the 5% level.<sup>21</sup>

In order to rule out the possibility that our main results concerning the treatment effect are driven by the higher frequency of coincidence of genders in the first pass, we disaggregate our data as shown in Table 3.

In line with the effect just mentioned, we find that, in the “coincidence” column,  $N_{01} > N_{10}$ , while the opposite holds for the “non-coincidence” column.<sup>22</sup> In the absence of an effect of requests for the receipt, we would

---

<sup>20</sup>This is true even when taking into account that it biases the reconstructed number of clients upwards.

<sup>21</sup>These figures are calculated using data from both passes.

<sup>22</sup>Notice that the gender of the vendor in the first and in the second pass is generally



	coincidence	non-coincidence
$N_{00}$	5	2
$N_{01}$	11	5
$N_{10}$	5	7
$N_{11}$	33	40

Table 3: Data disaggregated on coincidence of genders *in the first pass*.

expect the magnitude of the two differences to be the same: instead, in the “coincidence” case it is three times higher than in the other one. This difference in differences is precisely what is imputed to the treatment.

We compute the ATT also for different sub-samples. We find consistent results restricting the attention both to bakeries visited in the morning hours (54% of the sample) and to those visited in the afternoon: the ATT is always positive. Although it is higher in the morning (0.214) than in the afternoon (0.111), the difference is not statistically significant.<sup>23</sup> We find similar results when disaggregating on the (apparent) age of the vendor in the first pass:<sup>24</sup> the effect of the treatment is always positive, and no significant difference in its magnitude is found.

## 6 Conclusions

We estimated, through a field experiment, the causal effect of social pressure on tax compliance of shop sellers. In our experiment, social pressure takes the form of a request for the receipt, made to bakery sellers who do not release it. It implements the social norm of tax compliance in the relevant group of Italian citizens, manipulating the perception of the seller concerning Italian society’s “common view”. Our results support the established hypothesis according to which “*compliance cannot be explained entirely by the level of enforcement*” (Torgler (2002)), but rather it also depends on behavioral factors. In particular, we show that social pressure, even if expressed

---

unchanged (93.10% of cases): hence, if genders coincide in the first round they almost certainly differ in the second.

<sup>23</sup>We also observe that the rate of compliance on the first pass is homogeneous across hours of the day.

<sup>24</sup>Vendors were recorded as “young” when they were attributed 30 years or less (this measure has clearly no ambition of absolute precision).

through “cheap talk”, increases by 17.4% the propensity of sellers to release the receipt in the near future, and the result is significant at the 10% level. This finding also suggests the existence of a “social fiscal multiplier”: every request for a receipt causes the seller to release approximately 1.38 additional ones.

The effect of social pressure on the behavior of the seller could be explained in different ways. For instance, the seller may be *ashamed* of having received a rebuke. Alternatively, the seller may feel *embarrassed* by the discovery that he is acting *unjustly* (and possibly, that justice is more widespread than he used to think)<sup>25</sup>. These interpretations correspond to two main streams of literature explaining the high, and apparently “irrational”, fiscal compliance observed in advanced economies. The first motive is an example of *social stigma*, conceptualized by Gordon (1989) and others. The *embarrassment* for an *unjust* behavior calls instead for the *fairness-based* approach of Bordignon (1993). These two different approaches can be seen as corresponding to the concepts of *collective* and *subjective* moral constraints studied in laboratory experiments by Bosco and Mittone (1997).

Based on the structure of our experimental data, if the observed effect was related to *collective* moral constraints (*shame*) we would expect to find a higher impact of a rebuke when it is enacted in the presence of other clients. However such effect is not significant, possibly because our sample is not large enough to detect it. Our results are instead consistent with the idea that the intimate feeling of injustice plays a central role, an interpretation also supported by the experimental work of Bosco and Mittone (1997).

We also found that the probability of receiving a receipt is significantly lower when a client is of the same gender than the seller (−13.6%,  $\alpha = 5\%$ ). Since the gender of the agent was chosen independently of any characteristics of the bakery, the effect of the coincidence of genders has a causal interpretation. Moreover, its explanatory power is far higher than the effect of the mere gender of the client or of the vendor, which are non-significant (p-values of 30.4% and 62.21%, respectively).

This finding suggests that *illegality feeds out of complicity*, the latter being more scarce when individuals belong to different social groups (in this case, defined by gender). This finding provides additional evidence in favor of Torgler’s point of view presented above. A word of warning is however

---

<sup>25</sup>The seller could also expect that this goes hand in hand with an increase of fiscal controls on behalf of the authority.

required: the experiment involved *only one* agent of each gender. Additional evidence based on experiments involving more actors would be required to confirm that what we observe is indeed a consequence of the gender matching, rather than of individual characteristics of the agents. Moreover, studies conducted on a larger scale could analyze the effect of the coincidence of genders on the observed ATT.

The policy implications of our study consist in a strong support for awareness campaigns and other instruments aimed at influencing the behavior of sellers through the strengthening of social norms. In Italy, the sanctions on customers unable to exhibit a receipt after a purchase were abolished in 2003: reintroducing them could have a positive effect on fiscal compliance which goes beyond the *direct* fear of punishment.

Further research could be devoted to measuring the sensitivity of the results to changes in the location of the experiment (other cities), the type of shop, and the characteristics of the agents. Most importantly, experiments conducted with more than 2 agents acting consecutively may shed some light on the persistence of the effect of social pressure.

## References

- Allingham, M., Sandmo, A., 1972. Income tax evasion: A theoretical analysis. *Journal of public economics* 1 (3-4), 323–338.
- Alm, J., McClelland, G. H., Schulze, W. D., 1992. Why do people pay taxes? *Journal of Public Economics* 48 (1), 21–38.
- Alm, J., McClelland, G. H., Schulze, W. D., 2007. Changing the social norm of tax compliance by voting. *Kyklos* 52 (2), 141–171.
- Andreoni, J., Erard, B., Feinstein, J., 1998. Tax compliance. *Journal of economic literature*, 818–860.
- Bordignon, M., 1993. A fairness approach to income tax evasion. *Journal of Public Economics* 52 (3), 345–362.
- Bosco, L., Mittone, L., 1997. Tax evasion and moral constraints: some experimental evidence. *Kyklos* 50 (3), 297–324.
- Cowell, F. A., 1991. Tax-evasion experiments: an economist’s view. In: Webley (Ed.), *Tax Evasion: An Experimental Approach*. Cambridge University Press, pp. 123–127.
- Cummings, R. G., Martinez-Vazquez, J., McKee, M., 2001. Cross cultural comparisons of tax compliance behavior. *International Studies Program Working Paper Series*, at AYSPS, GSU.
- Cummings, R. G., Martinez-Vazquez, J., McKee, M., Torgler, B., Dec. 2006. Effects of tax morale on tax compliance: Experimental and survey evidence. *Working paper series*, Berkeley Olin Program in Law & Economics. URL <http://ideas.repec.org/p/cdl/oplwec/qt8sh2w9fp.html>
- Elffers, H., Weigel, R. H., Helsing, D. J., 1987. The consequences of different strategies for measuring tax evasion behavior. *Journal of Economic Psychology* 8 (3), 311–337.
- Erard, B., Feinstein, J. S., 1994. Honesty and evasion in the tax compliance game. *The RAND Journal of Economics* 5 (1), 1–19.

- Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature* 13 (1), 1–25.
- Fehr, E., Gächter, S., 1998. Reciprocity and economics: The economic implications of *Homo Reciprocans*. *European economic review* 42 (3), 845–859.
- Galbiati, R., Zanella, G., 2012. The tax evasion social multiplier: Evidence from Italy. *Journal of Public Economics* 96 (5), 485–494.
- Gordon, J., 1989. Individual morality and reputation costs as deterrents to tax evasion. *European Economic Review* 33 (4), 797–805.
- Graetz, M., Wilde, L., 1985. The economics of tax compliance: fact and fantasy. *National Tax Journal* 38, 355–363.
- Grasmick, H. G., Bursik Jr, R. J., 1990. Conscience, significant others, and rational choice: Extending the deterrence model. *Law and Society Review* 24 (3), 837–861.
- Groenland, E. A., Van Veldhoven, G. M., 1983. Tax evasion behavior: A psychological framework. *Journal of Economic Psychology* 3 (2), 129–144.
- Halla, M., 2012. Tax morale and compliance behavior: First evidence on a causal link. *The BE Journal of Economic Analysis & Policy* 12 (1).
- Kidder, L. H., Bellettirre, G., Cohn, E. S., 1977. Secret ambitions and public performances: The effects of anonymity on reward allocations made by men and women. *Journal of Experimental Social Psychology* 13 (1), 70–80.
- Kirchler, E., 2007. *The economic psychology of tax behaviour*. Cambridge University Press.
- Kleven, H., Knudsen, M., Kreiner, C., Pedersen, S., Saez, E., 2011. Unwilling or unable to cheat? evidence from a tax audit experiment in Denmark. *Econometrica* 79 (3), 651–692.
- Kolm, S.-C., 1973. A note on optimum tax evasion. *Journal of Public Economics* 2 (3), 265–270.

- Lewis, A., 2011. The social psychology of taxation. *British Journal of Social Psychology* 21 (2), 151–158.
- Paternoster, R., Saltzman, L. E., Waldo, G. P., Chiricos, T. G., 1983. Perceived risk and social control: Do sanctions really deter? *Law and Society Review* 17 (3), 457–479.
- Posner, E. A., 2000. Law and social norms: The case of tax compliance. *Virginia Law Review* 86 (8), 1781–1819.
- Reis, H. T., Gruzen, J., 1976. On mediating equity, equality, and self-interest: The role of self-presentation in social exchange. *Journal of Experimental Social Psychology* 12 (5), 487–503.
- Schwartz, R. D., Orleans, S., 1967. On legal sanctions. *The University of Chicago Law Review* 34 (2), 274–300.
- Sheskin, D., 2004. *Handbook of parametric and nonparametric statistical procedures*. CRC.
- Singh, B., 1973. Making honesty the best policy. *Journal of Public Economics* 2 (3), 257–263.
- Slemrod, J., 2007. Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives* 21 (1), 25–48.
- Slemrod, J., Blumenthal, M., Christian, C., 2001. Taxpayer response to an increased probability of audit: evidence from a controlled experiment in minnesota. *Journal of Public Economics* 79 (3), 455–483.
- Torgler, B., 2002. Speaking to theorists and searching for facts: Tax morale and tax compliance in experiments. *Journal of Economic Surveys* 16 (5), 657–683.
- Wärneryd, K.-E., Walerud, B., 1982. Taxes and economic behavior: Some interview data on tax evasion in sweden. *Journal of Economic Psychology* 2 (3), 187–211.
- Weigel, R. H., Hessing, D. J., Elffers, H., 1987. Tax evasion research: A critical appraisal and theoretical model. *Journal of Economic Psychology* 8 (2), 215–235.

- Wenzel, M., 2001. Misperceptions of social norms about tax compliance (2): A field-experiment. Centre for Tax System Integrity working paper.
- Wenzel, M., 2004. An analysis of norm processes in tax compliance. *Journal of Economic Psychology* 25 (2), 213–228.

## Appendix A Significance

Assuming  $\mathcal{H}_0 : ATT = 0$  holds, the number  $N_{01}$  is distributed according to a binomial where the number of draws is equal to the total number of treated bakeries, while the probability of each bakery changing compliance status by pure chance is  $p$ :

$$N_{01} \sim \mathcal{B}(p, N_0).$$

We ignore the true value of  $p$ , so we resort to the estimation  $\hat{p}$  presented above:

$$N_{01} \sim \mathcal{B}(0.17, 35).$$

Observing a value of  $N_{01} = 16$ , we can therefore calculate the probability of a type I error as:

$$\mathbb{P}\{N_{01} \geq 16 | N_{01} \sim \mathcal{B}(\hat{p}, N_0)\}.$$

Since the binomial probability distribution for  $k$  positive observations over  $n$  draws takes value  $\binom{n}{k} p^k (1-p)^{n-k}$ , the final estimate for the p-value is:

$$\sum_{k=N_{01}}^{N_0} \binom{N_0}{k} \hat{p}^k (1-\hat{p})^{N_0-k} = 0.0707$$

corresponding to the blue area in Figure 2.



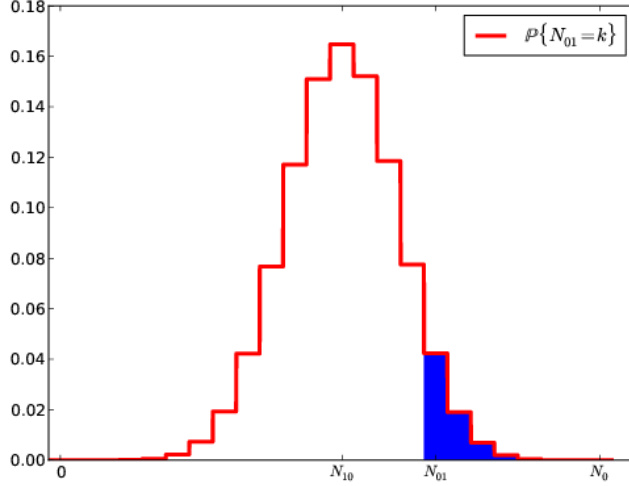


Figure 2: Probability distribution  $\mathcal{B}(\hat{p}, N_0)$ .

## Appendix B Studying the natural rate of alternation

We want to prove

$$\mathbb{E} \{c^2(0) | c^1 = 0\} = \mathbb{E} \{c^1 | c^2(0) = 0\}.$$

Considering that  $\mathbb{P} \{c^1 = 0\} = \mathbb{P} \{c^2 = 0\}$ , this easily reduces to proving that  $\mathbb{P} \{c^1 = 0, c^2 = 1\} = \mathbb{P} \{c^1 = 1, c^2 = 0\}$ .

In the absence of rebukes,  $c^1$  and  $c^2$  are functions of

1. the characteristics of the bakery (including both characteristics which are unchanged in time and characteristics which vary randomly - that means, *independently from the experiment*),
2. the characteristics of the agent buying the good,
3. the simple effect of time passing by,

4. the history itself of fiscal compliance (i.e.  $c^1 = 0$  may cause  $\mathbb{E}[c^2]$  to increase because of sense of guilt).

For what concerns point 1, characteristics which are constant in time clearly affect in the same way both  $c^1$  and  $c^2$ , while characteristics which vary in time but are independent from the experiment being in place clearly affect  $c^1$  and  $c^2$  *on average* in the same way. So they cannot be the cause of

$$\mathbb{P}\{c^1 = 0, c^2 = 1\} \neq \mathbb{P}\{c^1 = 1, c^2 = 0\}.$$

For what concerns point 2, the characteristics of the agents could very well affect  $c^0$  and  $c^1$  in different ways, except if the order of the agents was randomized - which is precisely what happens.

For what concerns point 3: we are assuming that the compliance propensity of a seller does not simply increase every 12 minutes - otherwise, it would change by huge amounts from the morning to the evening, and our data simply do not feature such an effect.

Now let's consider point 4. At a first sight, it would seem plausible that if, i.e., there is sense of guilt, we could find a higher probability for a "guilty" seller to repent ( $\mathbb{P}\{c^1 = 0, c^2 = 1\}$ ) than the opposite. For simplicity of exposition, let's discard for a moment the existence of points 1, 2, and 3.

Let us consider the following mental experiment: there are several agents, and each of them buys once a loaf of bread in the same bakery. They are spread out over the whole period of the experiment, entering one every 12 minutes during the opening hours. The result of (non-)compliance observations is a sequence which can be encoded as any of the possible strings of 0 and 1. Let us consider  $c^1$  a random component of this sequence, and  $c^2$  the following one.

- Among the possible sequences, is the one of the perfectly compliant seller:

$$\underbrace{1111111 \dots 1}_N :$$

it is clear that, in this case,

$$\mathbb{P}\{c^1 = 0, c^2 = 1\} = \mathbb{P}\{c^1 = 1, c^2 = 0\} = 0.$$

- If we change the above sequence by adding *a single* sale without the receipt being released, obtaining for instance

$$\underbrace{1111011 \dots 1}_N,$$

we now have a single possibility of observing a change in each direction:

$$\mathbb{P}\{c^1 = 0, c^2 = 1\} = \mathbb{P}\{c^1 = 1, c^2 = 0\} = \frac{1}{N}.$$

- Things don't change if instead than *a single* sale without receipt, we have *a group* of them:

$$\underbrace{1100011 \dots 1}_N.$$

- If we now add *another* block of sales without receipt, such as in

$$\underbrace{1100011 \dots 1001 \dots 1}_N,$$

the equivalence still holds, since

$$\mathbb{P}\{c^0 = 0, c^1 = 1\} = \mathbb{P}\{c^0 = 1, c^1 = 0\} = \frac{2}{N}.$$

The results clearly carries on, through induction, to any possible sequence of (non-)compliance: there will always be  $s$  changes from 0 to 1, and  $s$  in the opposite direction, for some  $s < N$ , so that

$$\mathbb{P}\{c^0 = 0, c^1 = 1\} = \mathbb{P}\{c^0 = 1, c^1 = 0\} = \frac{s}{N}.$$

The only case in which this is not true is if the receipt was omitted at the *first* sale, or the *last* (but *not both*): in this case, there will be an unmatched “01” without the corresponding “10”:

$$\mathbb{P}\{c^0 = 0, c^1 = 1\} = \frac{s}{N} \neq \mathbb{P}\{c^0 = 1, c^1 = 0\} = \frac{s-1}{N}$$

(or vice-versa). This difference is however negligible for high enough values of  $N$  (which is, in our case, of order of magnitude of the hundreds).

## Appendix C Reconstructing the DiD

We can trace back our experiment to a Difference in Differences model. We do not have a unified time break; instead, for every bakery we have a “before” ( $c^1$ ) and an “after” ( $c^2$ ). In the meanwhile, some bakeries are treated - that is, rebuked.

Let us start from the general form for the DiD:

$$\begin{aligned} ATT &= (\mathbb{E}\{c^2|D = 1\} - \mathbb{E}\{c^1|D = 1\}) \\ &\quad - (\mathbb{E}\{c^2|D = 0\} - \mathbb{E}\{c^1|D = 0\}) \end{aligned} \tag{C.1}$$

Notice that, given two bakeries characterized by the same (unobservable) level of compliance, if the first releases the receipt at some time but the other does not, then the temporal evolution of fiscal compliance *will be different*, in the sense that the first can only become less compliant, while the second can only become more compliant.

This means that, for the Difference in Difference method to give unbiased results, we must imagine that also all the untreated bakeries refrained from releasing the receipt on the first pass.

Under this assumption, we have  $\mathbb{E}\{c^1|D = 0\} = 0$ .

We instead lack an observational value for  $\mathbb{E}\{c^2|D = 0\}$ ; we can however resort to exploiting assumption (4), according to which  $\mathbb{E}\{c^2(0)|D = 0\} = p$ .

Given these premises, (C.1) will be evaluated as

$$\begin{aligned} ATT &\approx \left( \frac{N_{01}}{N_0} - 0 \right) - (\hat{p} - 0) \\ &= \frac{N_{01}}{N_0} - \hat{p} \end{aligned}$$

as above.

In general, the *selection bias* could be calculated as

$$\mathbb{E}\{c^1|D = 1\} - \mathbb{E}\{c^1|D = 0\}$$

which in our data would correspond to  $0 - 1 = -1$ . However, it is important to remark that this number sums up two different effects:

1. bakeries being treated have, on average, a lower *intrinsic* propensity to release receipts,

2. on top of that, among *occasionally* unlawful bakeries, the one being treated are “caught” in a *contingent* “non-complying moment”.

Effect 1 is the one taken into consideration when deriving the ATT, since effect 2 does not affect the results from the second pass in any way.