

RANDOMIZED TAX ENFORCEMENT MESSAGES: A POLICY TOOL FOR IMPROVING AUDIT STRATEGIES

Dina Pomeranz, Cristóbal Marshall and Pamela Castellón



SYNOPSIS

Reducing tax evasion is a key challenge for governments around the world, particularly in developing countries. This paper presents a methodology to generate information to optimize audit strategies. Randomly selected taxpayers receive a deterrence message. Comparing their subsequent tax payments to a control group allows estimating what types of taxpayers are more likely to respond to an increase in perceived audit probability. This information can be used to target audits toward taxpayers that respond particularly strongly, and to construct risk indicators to predict taxpayers' responses. We show results from an application in Chile and describe lessons learned during the implementation.

***The Authors:** Dina Pomeranz has a PhD in Economics from Harvard University and is an Assistant Professor at the Harvard Business School. She is a faculty research fellow at the National Bureau of Economic Research (NBER) and an affiliate professor at the Abdul Latif Jameel Poverty Action Lab (J-PAL). Cristobal Marshall has a B.A. in Economics from the Catholic University of Chile and a Master's degree in Public Administration in International Development from the Harvard Kennedy School. Pamela Castellón is an industrial engineer and has a master's degree in Operations Management from the University of Chile. She is a professional in the Department of Economic and Taxation Studies at the Internal Revenue Service of Chile (SII).*

Contents

1. Methodology
2. Implementation: The case of Chile
3. Implementation lessons learned
4. Cost-benefit analysis
5. Conclusions
6. References
7. Appendix

Conducting successful and cost-effective audits is no easy task. Audits are costly for both tax administrators and taxpayers, and tax administrations often have limited information and resources to design and optimize their audit strategies. Audit performance is often reported to be a particularly weak element of tax administration (Ebrill et al., 2001).

In this paper, we present a methodology for a critical component of an audit strategy: the selection of taxpayers to be audited.¹ Risk-based selection methods are increasingly used by tax administrators as a tool to allocate their audit activities. The goal of these methods is to identify those taxpayers who are most likely to respond strongly to an increase in the audit probability. Audits are then conducted based on that assessment, rather than randomly or on an ad-hoc basis. Various analytical prediction techniques such as data matching, data mining, and parametric and non-parametric models are used to identify high-risk cases.² A score can then be assigned to each individual or business based on their identified characteristics. This is similar, for example, to risk indicators created in the financial industry for credit risks. For the construction of such risk indicators, we also

present a way to overcome the methodological challenge of combining data obtained from randomized interventions with data-mining-based prediction models.

A key challenge when using these methods, however, is to accurately identify what information to use as input into the risk prediction models. For that purpose, we present a methodology to improve the allocation of audits using randomized deterrence messages. This approach avoids both the high costs of randomized audits, and the problems of selection bias inherent in using information based on non-randomized past audit histories as input for the prediction models. We illustrate our method based on a randomized impact evaluation conducted by Pomeranz (2013) in collaboration with the Chilean Tax Authority, Servicio de Impuestos Internos (SII). This approach can be implemented in many different settings and for different types of taxes, at a very low cost compared to the gains in revenue. Furthermore, a similar strategy can also be used in other contexts, for example in the case of organizations that use audits or similar monitoring processes as part of their operations.

Developing a risk-based model or a scoring technique faces two main challenges. First, the variable of interest is not easily observable. For some risk-based indicators in other contexts, such as credit scoring, the outcome of interest – credit default – is easily observable. Financial institutions have direct access to the information of credit default for all their clients, which, when combined with clients' characteristics, can be used to develop a scoring model that predicts creditworthiness. Tax non-compliance, on the other hand, similar to corruption or crime activities, is an act of concealing information. Collecting data on this variable requires measurement efforts which, as in the case of audits, are often very costly.

-
1. *This project would not have been possible without the generous support of the following institutions: Chilean Tax Authority - Servicio de Impuestos Internos (SII), the Center for International Development at the Harvard Kennedy School, the David Rockefeller Center for Latin American Studies, the Project on Justice Welfare and Economics at Harvard University and the Swiss Study Foundation. All errors are our own, and we are available for any questions about the implementation of the study and the potential application of the method in other contexts.*
 2. *For a detailed description of risk-based methodologies see Vellutini (2011).*

Second, when data exists, for example from prior audits, it is often difficult to gather conclusive insights because of selection bias. Audits are usually not randomly conducted. An analysis that compares data from audited taxpayers with that from non-audited taxpayers faces such selection problems, since taxpayers in the two groups are not necessarily comparable. Firms that have been selected for audits in the past are different from those that have not, which can result in erroneous interpretations about what type of taxpayer has the highest risk. A high-quality prediction method therefore requires an identification strategy that can determine a causal relation between compliance and a set of taxpayer characteristics. However, most risk-based analysis on taxes lacks external variation in a variable that can potentially affect compliance.

Our methodology addresses both of these challenges. We use a randomized control trial in which the perceived audit probability of randomly selected taxpayers is increased through a notice from the tax authority informing them that they are under special scrutiny. The response to the notice is measured by comparing the subsequent tax payments of those randomly assigned to receive the notice, to those in the control group, who did not receive the notice. Furthermore, analyzing the differences in the responses allows us to investigate what types of taxpayer characteristics are associated with a stronger response. These characteristics, in turn, can be fed into the risk indicator calculation, which helps identify the most optimal candidates for an audit. The goal of this study is to target resources where the expected increase in tax payments is the highest.

Our proposed methodology also addresses a common concern related to the implementation costs of randomized audits. Randomized deterrence messages can capture similar information to what would have been obtained by conducting an actual audit, but at a fraction of the cost. Sending a deterrence letter from the tax authority measures the response of

a taxpayer to the perceived probability of being audited. The letter thereby increases the expected cost of non-compliance, thus increasing the incentive to declare more taxes. As taxpayers may also vary in how they react to the threat of an audit, this strategy does not necessarily capture taxpayers who are evading the most, but rather those who respond the most to an increase in the probability of being audited. The results are thus aligned with strategy of increasing revenues. In addition, the low cost associated with sending the letters allows authorities to work with data for a large number of individuals, thereby increasing the precision of the analysis.

The issue of selection bias is addressed by sending the letters in a randomly selected manner. We can thus obtain an unbiased estimation of taxpayer responses to an increase in the threat of being audited, and of how different taxpayer characteristics are associated with stronger responses. Once that information is collected, we construct a very simple risk indicator for each taxpayer that provides the probability of responding to the higher threat. In practice, more complex risk indicators can also be employed.

The methodology consists of three steps:

1. **Deterrence notice:** The first and most involved step consists of a randomized controlled trial in which the perceived audit probability of randomly selected taxpayers is increased through a notice from the tax authority informing them that they are under special scrutiny.
2. **Estimating the reaction to the notice:** The response to the notice can be measured by comparing the subsequent tax payments of those randomly assigned to receive the notice to those in the control group, who did not receive any notice.
3. **Extracting information for improved audit targeting:** Comparing the different responses allows tax administrators to investigate what types of taxpayer characteristics are associated with a stronger response to the

increase in the perceived audit probability. This information can also be summarized in an overall risk indicator that estimates for each taxpayer, the likelihood of responding positively to an increased audit probability.

The remainder of this paper is organized as follows: Section 1 describes the methodology

and the steps involved in its implementation. Section 2 provides background information and basic results from the intervention in Chile. Section 3 outlines some of the lessons learned during this implementation. Section 4 discusses the costs and benefits associated with the implementation of this methodology, and Section 5 concludes.

1. METHODOLOGY

This section describes the three steps involved in the process of setting up and analyzing randomized deterrence message to optimize tax enforcement.

1.1. Implementation of the deterrence notice

The first step toward estimating taxpayers' responses to an increase in perceived audit probability requires a credible strategy to change their expectations. This can be achieved through an official message from the tax authority informing them that they have been selected for special scrutiny. The credibility of the deterrence message is of great importance, both to ensure the effectiveness of this strategy to gain information to optimize tax enforcement, and to avoid undermining the credibility of the tax authority in the process of this intervention (the latter will be discussed in more detail in Section 4).

The implementation of randomized deterrence messages involves a number of key steps that lead to a valid analysis:³

- **Choice of the universe of taxpayers to be studied:** The process should start by identifying the universe of taxpayers for which information about the response to audits will be analyzed. In the selection of the relevant

universe, tax authorities should consider their capacity and limitations in order to tailor their audit efforts based on the information that will be obtained. For example, the analysis could be based on all individuals subject to the personal income tax, or all independent entrepreneurs, etc. However, it is important that the universe be large enough to allow for the detection of statistically significant effects (see below).

- **Choice of the content of the deterrence notice:** Once the target universe has been selected, the content of the deterrence message can be developed. For practical reasons, two issues should be kept in mind when choosing the wording of the message. First, the message should credibly and effectively communicate an increase in the perceived probability of being audited. Second, the message should be straightforward and minimize possible confusion on the part of the recipient in order to avoid an influx of inquiries, concerns, or complaints to the tax authority's helpline or offices. It is advisable to conduct a series of qualitative pre-tests and informal interviews with taxpayers in order to ensure that the message has the intended effect and that the wording is clear to a broad range of taxpayers with varying backgrounds.

3. These steps usually also apply to the implementation of randomized evaluations in many other areas and contexts.

- Choice of the number of taxpayers to receive the deterrence notice:** To be able to detect a statistically significant impact of the deterrence message and gather information on differences between different types of taxpayers, the sample size in both the treatment and the control groups needs to be sufficiently large. The method used to determine the minimum number of observations required for statistical significance is called power calculations.⁴ Power calculations should always be performed as a key initial step. In general, the higher the number of characteristics and subgroups for which one would like to estimate differential effects, the larger the required sample size. In other words, the more fine-tuned the analysis is intended to be, the larger the required number of taxpayers in the analysis. Note that the number of taxpayers in the treatment and control groups need not necessarily be the same. In fact, it is advisable that all taxpayers who were randomly selected not to receive the notice be included in the control group, even if the number of individuals in this group is many times larger than the number in the treatment group.
- Random selection of those taxpayers slated to receive the deterrence notice:** It is crucial for the validity of the analysis that taxpayers be selected in a random process from among the universe selected for the analysis. The randomization creates two groups of taxpayers that are identical in all observable and non-observable characteristics, except for the receipt of the deterrence notice. All differences in terms of subsequent tax payments between the treatment and the control group can therefore be attributed to the deterrence notice.⁵ For this to remain true, it is important that subsequent to the random assignment, taxpayers in both the treatment and control groups continue to be treated exactly the same in all aspects except for the deterrence notice. Also, the control group has to consist only of taxpayers that were randomly selected not to receive the message. Any taxpayers who were not part of the original universe for which the randomization was performed cannot be included in the analysis.
- Pilot:** Before the mass mailing or distribution of the deterrence notice, it is highly advisable to conduct a pilot of the intervention with a small sub-sample in order to verify whether all the required actions can be implemented as planned. Typically, important lessons are learned during the pilot phase and unexpected obstacles can be discovered and removed before the large scale distribution is implemented. Pilots can therefore contribute significantly to the successful implementation of the proposed methodology.
- Ensuring that the random selection is respected in the implementation:** For the validity and success of the analysis, it is important that the original random assignment be preserved and not tampered with. This means two things in practice: first, ensuring that as much as possible, those taxpayers that have been selected to receive the notice indeed receive one, and those, who have been selected to be in the control group, do not receive it (see discussion in Section 4 below on why that might not always be possible); second, where this is not fully possible, taking care to maintain the information about who was originally assigned to the treatment and control groups. In order to ensure the validity of the analysis, it is imperative that all estimation be conducted based on this original random assignment.

4. Power calculation can be applied through programs such as “Optimal Design” <http://www.wtgrantfdn.org/resources/research-tools>. For more details on power calculation see also Duflo et al. (2007).

5. For more details on the rationale and benefits of using randomization see Duflo et al. (2007).

1.2 Estimating the reaction to the notice

The second step of the methodology is to estimate the magnitude of taxpayers' reactions to the higher expected probability of being audited. To obtain an unbiased estimate of this reaction, we rely on the randomized assignment of the notice, as discussed above. Due to this random assignment, the effect of the deterrence notice can be measured by simply comparing the tax payments in the period following the distribution of the notice, of those who were selected to receive the notice, with those in the control group.⁶

The overall impact of the message can be estimated using regression analysis, following the specification shown in Equation (1). This specification is a difference-in-difference approach. A simple difference of the post-treatment outcomes would also be valid. The inclusion of the pre-treatment period and the taxpayer fixed effect simply improves the precision of the analysis.

$$y_{it} = \alpha + \beta_1 tZ_i + \beta_2 Z_i + \lambda_i + \delta_t + e_{it} \quad (1)$$

In regression (1), y_{it} represents the outcome variable of interest for taxpayer i at time t , and Z_i is a dummy variable that indicates whether the taxpayer has been assigned to the treatment group. tZ_i indicates whether the taxpayer belongs to the treatment group, and the time is in the post-treatment period. δ_t stands for a year or month fixed effect, λ_i is a taxpayer fixed effect, and e_{it} is an error term. This regression estimates the coefficient of interest, β_1 , which represents the overall treatment effect. It indicates the difference in the outcome of interest caused by the treatment.

One choice of variables is the outcome of interest. A useful metric in this regard is the amount of tax declared. For estimation purposes, different

specifications of this variable can be employed, including the following:

- **The mean amount of taxes declared:** This basic specification provides an easy interpretation of the estimated effects. It indicates the additional amount of taxes paid among those who received the deterrence notice. It is therefore also informative for cost-benefit analyses of different audit strategies. At the same time, this specification has two potential drawbacks. First, it often leads to relatively low statistical power, and it therefore provides less precision to estimate differential responses by different types of taxpayers. The specific amounts can be driven by outliers, and may therefore sometimes be misleading. A second drawback relates to the interest to study not only which taxpayers respond most in terms of absolute financial amounts, but also in relative terms. For example, an increase in \$100 for a small store is very different from an increase of the same amount by a large factory. However, the magnitudes of the estimations of the mean are driven by the relative size of taxpayers. This can make them less useful when interested in estimating differential behavior by type of taxpayer (e.g. comparing small vs. medium size firms).
- **The log of the amount of taxes declared:** Since the log approximates changes of the outcome to changes in percentage terms, it can be useful to estimate relative changes and therefore gives a better indication of the behavior change of different taxpayers. However, logs cannot be applied if the outcome variable includes negative amounts, such as tax obligations below zero that can subsequently be employed as carry-over tax credits in the following tax declaration. This is for example the case in the Chilean VAT.

6. Note that the validity of the analysis is compromised if it compares those who received the message to taxpayers who were not included in the universe for which the randomization was conducted. Instead, the analysis must compare all those assigned to the treatment group to all those assigned to the control group. The original assignment needs to be used whatever be their actual treatment status (intent-to-treat, ITT). If the relevant information is the effect of the intervention itself rather than the original assignment, then the effect of the treatment-on-the-treated (TOT) can be used. See Duflo et al. (2007) for more details on ITT and TOT.

- **The median of declared taxes:** A quantile regression is a specification that can also be used when there is large dispersion in the data, since the median is less sensitive to extreme values than the mean. It has the benefit of still being applicable when there are negative values in the data. However, similar to the mean, the magnitude of the estimates is driven by the relative size of the taxpayers.
- **The probability of declaring positive taxes:** In many cases, a margin of interest can be whether taxpayers are declaring any taxes at all, either by submitting a declaration or not, or if a declaration is submitted, declaring a positive amount. The impact of the deterrence notice on this behavior can be measured by a linear probability regression. This is a regular OLS regression, where a dummy variable indicating whether the taxpayer declares positive taxes is the outcome variable.
- **The probability of declaring higher taxes compared to last year:** Another specification that can be very useful is a linear probability regression, where the outcome variable is a dummy indicating whether declared taxes are higher than in the previous year. This specification is robust to outliers and large variances in the data, is not driven by the taxpayer's size, and can deal with negative amounts of the tax variable.

The choice of the outcome variable to be used for estimation purposes can be made on a case-by-case basis, depending on the data available, the distribution of the data, and the type of analysis the tax authority is interested in. In Section 3, we discuss the variable chosen for our implementation in Chile and the reason behind that choice.

How should the responses to the letter be interpreted? Can the magnitude of the taxpayers'

reactions be used as a good estimation of overall tax evasion? Not necessarily. While paying higher taxes in response to an increased audit probability suggests that taxpayers might have been underreporting tax liabilities before receiving the deterrence notice, the amount of increased payments may not be a good estimator of the amount of taxes evaded. Taxpayers may still engage in evasion after receiving the notice if, for example, they are confident that their evasion cannot be detected, if they expect the fine to be low or non-enforced, or if they doubt the credibility of the deterrence message. On the other hand, some taxpayers may respond to the audit notice by over-declaring taxes. Therefore, the response to the letter should not be interpreted as an accurate estimate of tax evasion but instead as the response to a higher expected probability of being audited.

1.3 Extracting information for improved audit targeting

1.3.1 What types of taxpayers respond more?

The goal of this approach is not merely to estimate the magnitude of taxpayers' overall reactions to the message, but also the kinds of taxpayer characteristics that are associated with a stronger reaction to the notice, such as firm size, industry, geographical location, or past behavior.

Similar to an actual audit, the response of taxpayers to a deterrence notice will vary. The variables that can influence the magnitude of taxpayers' response to the threat of audits include: (i) the amount of tax currently being evaded, (ii) the expected audit probability prior to receiving the notice of increased scrutiny, (iii) the expected penalty and the extent to which it is anticipated to be actually enforced, (iv) the expected likelihood that an audit will reveal evasion, and (v) the taxpayer's degree of risk aversion, to name a few.

For the estimation of the differential response by type of taxpayer, we use a variation of regression (1), which is captured by the following specification:

$$y_{it} = \alpha + \beta_1 tZ_i + \beta_2 tX_i + \beta_3 tZ_i X_i + \delta_t + \lambda_i + e_{it}, \quad (2)$$

where X_i is a vector of taxpayer characteristics, and $tZ_i X_i$ is an interaction term of taxpayer characteristics X_i and the post-treatment indicator described above. From this regression, we can estimate the coefficient of interest, $\hat{\beta}_3$, which represents the interactions of the treatment effect with taxpayer-specific characteristics.⁷ For example, if the taxpayers that are analyzed are firms, we can classify them by level of sales according to five categories: no sales, micro size, small, medium or large. Then, X_m, X_s, X_M, X_L represent dummy variables indicating the respective firm sizes (“no sales” is the omitted category). In this case, the estimated coefficient $\hat{\beta}_3$ for the $tZ_i X_m$ variable would show the differential impact of the deterrence letter for micro size firms compared to the effect on firms that had no sales. A positive estimated coefficient means that, all else equal, micro size firms react comparatively more than those with no sales. If the goal of a tax authority is to increase tax revenues, then audits should target micro firms more heavily than firms with no sales. This is precisely what will be captured by our proposed risk indicator.

1.3.2 Constructing a risk indicator

Rather than analyzing the differential effects for specific taxpayer characteristics in isolation, one can also aggregate the information on how taxpayer characteristics interact with the treatment into one single risk indicator. One

very simple way in which this can be done is by using the estimated coefficients for β_1 and the differential effects by taxpayers’ types β_3 from regression (2). Together, these predict the response to the deterrence notice by all taxpayers based on their characteristics. This information is often available and continually collected by most tax authorities, allowing a regular calculation of the risk indicator for all taxpayers.

The risk indicator is therefore calculated as follows:

$$r_i = \hat{\beta}_1 + \hat{\beta}_3, \quad (3)$$

where r_i provides a numerical value for each taxpayer, indicating the expected response to an increase in the audit probability. This risk indicator can be directly applied to the design of an audit strategy. A tax authority could, for example, rank taxpayers in term of r_i and then audit those with the highest values of the indicator.

It is worth noting that in the case that the independent variable is defined as the probability of paying positive taxes or higher taxes than the previous year, r_i is independent of the relative size of taxes paid, and higher values of r_i do not necessarily represent higher additional revenues. If a tax authority wants to target taxpayers with a higher probability of responding to the deterrence notice as a way to target those taxpayers most suspicious of tax evasion, then r_i is the appropriate indicator to use. If instead the tax authority is interested in maximizing tax revenues, it might want to multiply r_i by a variable that captures taxpayers’ size, such as past revenues or taxable income, and use this new value as the risk indicator.

7. If the specification employed is not a panel regression and therefore does not include taxpayer fixed effects, X_i needs to be included on its own in the regression as well, in place of λ_i .

2. IMPLEMENTATION: THE CASE OF CHILE

This methodology has been implemented in close collaboration with the Chilean Tax Authority, Servicio de Impuestos Internos de Chile (SII). SII has the reputation of being highly effective and is one of the most respected institutions in the country (Adimark, 2006). This reputation has been obtained over time through various factors, such as the provision of high-quality public service, constant service improvement, and application of cutting-edge methods and technologies. Following this tradition, SII decided to implement the method of randomized deterrence notices as part of a continuous effort to improve its processes and identify new mechanisms in order to make tax collection and administration more effective.

In the following, we discuss how each of the implementation steps laid out in Section 2 was put into practice in Chile.

2.1 Implementation of the deterrence notice

- Choice of the universe of taxpayers to be studied:** The analysis in Chile focuses on the Value Added Tax (VAT), which represents approximately half of tax revenues of the country (Servicio de Impuestos Internos, 2012). Although the VAT is believed to have distinctive features that make it less vulnerable to evasion and fraud, in recent years there has been an increasing concern regarding VAT revenue losses associated with non-compliance in both developed and developing countries (Keen and Lockwood, 2010). The universe of taxpayers included in the Chilean analysis consisted of all small and medium size firms in the country that were operating in June 2008, had declared a positive amount of VAT for at least one
- Choice of the content of the deterrence notice:** With the aim of increasing taxpayers' perceived audit probabilities, SII sent deterrence letters to a randomly selected group of small and medium size firms.⁸ In particular, the letters informed taxpayers that in a random selection process among micro, small and medium size firms, they had been selected for analysis, and in case of any detected irregularities, they could be summoned to an audit. In order to ensure that the message had the intended impact, a series of informal interviews were conducted with firm representatives prior to the mailing. During these interviews, owners or accountants of small and medium size firms were asked how they would interpret the message and what they would do if they received one. These pre-tests helped fine-tune the message. They also alerted the tax authority as to the usefulness of mentioning in the letter that taxpayers did not need to take any immediate action or come to the tax authority office to inquire for further information. The content of the messages was also vetted by the internal legal and communications departments of SII.
- Choice of the number of taxpayers to receive the deterrence notice:** Power calculations showed that among the selected sample of 445,734 taxpayers in Chile, mailing of the deterrence message to 102,000 was required to have a high enough probability

8. To test whether the impact of the letter was due to the deterrence message or due to the simple fact of receiving mail from the tax authority, motivational and placebo letters that contained information irrelevant for tax compliance, were also sent to randomly selected taxpayers. For more details see (Pomeranz, 2013).

to obtain statistically significant results. As described in Section 2, treatment and control groups do not necessarily have to have the same number of units. Since the control group in this case has no cost, and since the notice might lose credibility if too large a share of firms receives it, it can often result that the treatment group is smaller than the control group in this type of intervention.

- **Random selection of the taxpayers slated to receive the deterrence notice:**

The random selection of taxpayers was conducted electronically in order to avoid any possible human interference and therefore ensure the validity of the process. To increase the precision of the randomization process at making treatment and control groups comparable in all characteristics, the sample was stratified by a few key firm characteristics, such as size, region, past audits, etc. Randomization then took place within these subgroups.

- **Pilot:** A pilot of 250 letters was sent out two months prior to the large-scale mailing, helping detect wrinkles in the process. It allowed adjusting several internal procedures, as well as the wording of the letters, and ensured a smooth implementation of the large-scale mailing.
- **Ensuring that the random selection is respected in the implementation:** The letter messages were mailed in envelopes by the VAT department of the tax administration.

The notice was sent using certified mail, so that information could be obtained on which taxpayers actually received the letter and by what date. In order to minimize the cases where taxpayers in the treatment group received no letter, taxpayers with invalid postal addresses were excluded from the universe. It was very important to exclude those with invalid mailing addresses both from the treatment and the control group. If they had been excluded only from the treatment group, this would have rendered the comparison between the remaining treatment group and the control group invalid. The exclusion of those with invalid addresses allowed for a substantial increase in the statistical power of the experiment by significantly reducing non-compliance with treatment assignment.⁹

2.2 Estimating the reaction to the notice

2.2.1 Summary statistics at baseline

The first step of the analysis in any randomized control trial is to verify that the randomization has been done properly, and that the treatment and control groups are not statistically significantly different from each other prior to the intervention. Table 1 presents such baseline summary statistics for the taxpayers in the treatment group and the difference with respect to the control group. As one would expect given the random assignment, average characteristics between the two groups look very similar, and none of the differences are statistically significant.

9. *Partial compliance (i.e. if not all those assigned to treatment actually receive treatment, or if some assigned to the control group still receive the treatment) strongly reduces the statistical power of randomized evaluations. The higher the non-compliance rate, the higher the sample size that is required to estimate a minimum effect. To estimate the direct effect on the compliers, a treatment-on-the-treated (TOT) estimation needs to be applied in these cases.*

Table 1
Baseline Summary Statistics and Balance of Randomization

	Control Group	Difference of the Treatment Group
Monthly tax paid (mean)	264,434 (2,746)	1,342 (10,144)
Monthly tax paid (median)	69,892 (458)	-779 (920)
Taxpayer age in months	108 (0.12)	-0.32 (0.25)
Non-filed declarations	4.3 (0.02)	-0.03 (0.05)
No sales year prior	1.5 (0.02)	-0.05 (0.04)
Micro size	74.5 (0.08)	-0.04 (0.16)
Small size	18.2 (0.07)	0.03 (0.14)
Retail taxpayers	2.8 (0.03)	0.01 (0.06)
Intermediary taxpayers	28.7 (0.08)	-0.06 (0.16)
Intermediary taxpayers	38.2 (0.09)	0.15 (0.18)
Final sales	45.6 (0.08)	-0.10 (0.17)
Number of taxpayers	306,605	102,031

Notas: Cada fila muestra el resultado obtenido por regresión de la variable de interés en cuestión con un término constante y una variable dummy que indica el grupo al que pertenece. El término constante captura el valor que correspondiente al grupo de control. Los errores estándar robustos (entre paréntesis) se agrupan en el nivel de los contribuyentes. Los importes monetarios son en pesos chilenos (500 pesos son aproximadamente equivalentes a 1 USD). Ninguna de las diferencias es estadísticamente significativa en un nivel de significancia del 10%. Todos los datos (incluyendo las características de los contribuyentes) se obtuvieron de los registros oficiales de impuestos.

The first two rows of Table 1 show whether taxpayers are balanced in terms of their tax payments prior to sending the notice. The group of firms that took part in our analysis paid an average of 264,000 pesos (about 500 USD) in monthly VAT, with a median of approximately 70,000 (about 140 USD). The large difference between the mean and the median indicates a very large dispersion in the distribution of tax payments. As discussed below, this large dispersion has implications for the regression specification used for the empirical analysis.

Next, we look at potential differences in terms of size. There are five official firm size categories in Chile, based on the amount of sales in the previous tax year: firms with no sales, micro-sized, small-sized, medium-sized and large-sized firms.¹⁰ The micro-sized taxpayers are by far the largest group, comprising 74.5% of the universe. The second largest group is the small-sized taxpayers (18.2%), followed by the medium-sized taxpayers (2.8%) and taxpayers with no sales in the preceding year (1.5%). The remaining 3% are those new taxpayers that have not yet been classified. Taxpayers in treatment and control groups are balanced in terms of the proportion of firms of each size.

Taxpayers are also balanced across treatment groups with respect to their position in the production chain: 28.8% are retailers that sell only to final consumers, and 38.2% are intermediary taxpayers that sell only to other taxpayers. Overall, the share of sales to the final consumer is 45.6%. Finally, the treatment and the control groups are similar in terms of age. On average, firms started operating approximately 9 years before the baseline data was collected.

2.2.2 Overall impact of the deterrence letter

Once treatment-control balance has been verified, we estimate the effect of the deterrence letter using regression (1). Given the large dispersion in the distribution of declared taxes, analyzing the impact of the intervention on the mean of declared VAT does not lead to statistically significant results, as the variance is extremely large. Thus, we estimate a linear probability version of regression (1), where the outcome variable is the probability of paying higher VAT after receiving the notice, compared

to the same month of the previous year. This specification has several benefits compared to alternative measures. First, as opposed to a log specification, it can be applied to variables that include zero or negative values, which is the case for declared taxes.¹¹ Second, it is robust to outliers and high variances, as is the case with our fat-tailed distribution of declared taxes. Third, it takes into account firms in all parts of the distribution of VAT payments (as opposed to measures such as the probability of declaring positive taxes, the mean, or the median, which are all sensitive to taxpayers' size), making it useful to compare the treatment effect between different types of firms.¹²

The results show a significant impact of the deterrence message. The estimated overall treatment effect is a 1.40 percentage point increase in the probability of paying higher VAT than the same month of the previous year. The result is statistically significant at the 1% level (Pomeranz, 2013). Two other interesting results from the evaluation, also shown in Pomeranz (2013), are worth noting. First, the impact in terms of higher taxes seems to be very persistent. The initial impact in the treatment group is immediate and decreases only slowly over the following months. It reaches the same level as the control group again after about 18 months. In the first six months of the implementation of the enforcement letter, the increase in tax revenue is estimated to be about 3.3 million USD. Second, increasing the audit probability of firms generated spillovers up the VAT paper trail that lead to an increase of their suppliers' tax payments. This result suggests that when designing audit strategies, tax authorities should not only consider the effect on the audited firm, but also the spillover effects on the firm's trading network.

10. Large firms are not included in this study.

11. In Chile, firms can declare negative VAT amounts. In this case, they pay zero VAT for the corresponding month, and carry the negative tax amount over to their next declaration in form of a tax credit.

12. See Pomeranz (2013) for a detailed discussion on the use of alternative dependent variables, and why this particular outcome variable was chosen in the Chilean case.

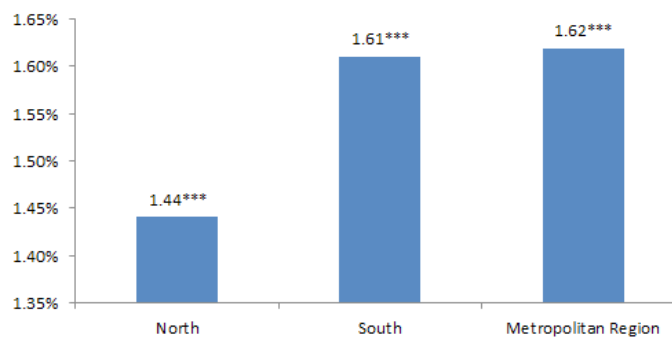
2.3 Extracting information for improved audit targeting

2.3.1 What types of taxpayers respond more?

Now that we have analyzed the overall impact of the deterrence notice, we can start looking at differential impacts by different types of taxpayers. We start by dividing the taxpayers according to three regional groups - North, South,

and Metropolitan Region - and estimate the effect of the notice for each subgroup. Figure 1 shows that the estimated effect of the deterrence letter is higher in the Metropolitan Region and in the South and lower in the North.

Graph 1
Impact of the Notice by Region



Notes: Each bar represents a separate regression for each region. The numbers on top of the bars indicate the coefficient on being in the treatment group. The height of the bar represents the effect of receiving the notice as a percentage point increase in the probability of paying higher taxes than last year. *** = significant at the 1% level, ** = significant at the 5% level, * = significant at the 10% level.

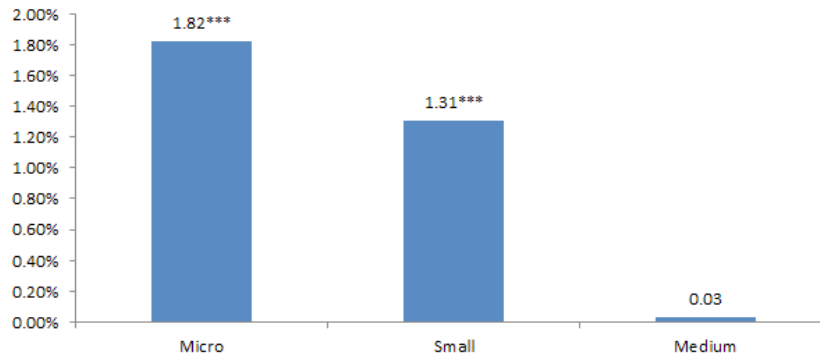
The differences in response by region may partly be due to the fact that firm characteristics vary by region along dimensions that may potentially influence the degree of response to the deterrence notice, such as firm size, economic sector, or the share of sales that goes to final consumers. The following figures present the differential effects along these dimensions.

Figure 2 shows the impact of the notice by firm size. The response is highest among firms of smaller size, which is consistent with the claim that smaller taxpayers evade more taxes (Kleven et al., 2009). Internal processes that

keep track of transactions, external monitoring mechanisms, and reputational effects are more likely to be present among larger taxpayers, making them less likely to respond to an additional increase in the audit probability

(Kopczuk and Slemrod, 2006). According to these results, micro-size firms that received the letter had a 1.8 percentage point higher chance of increasing their VAT payments since the previous year. The effect is 1.3 percentage points among small taxpayers and there is almost no effect among medium size taxpayers.

Graph 2
Impact of the Notice by Firm Size

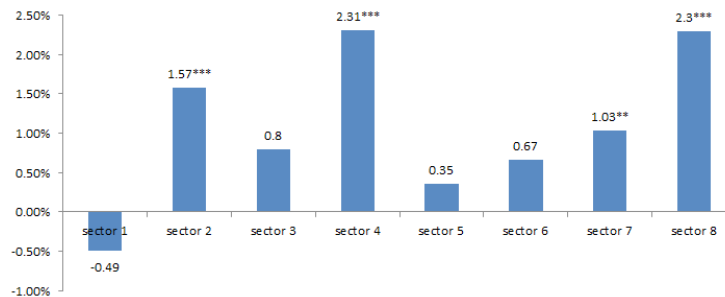


Notes: Each bar represents a separate regression for each size category. The numbers on top of the bars indicate the coefficient on being in the treatment group. The height of the bar represents the effect of receiving the notice as a percentage point increase in the probability of paying higher taxes than last year. *** = significant at the 1% level, ** = significant at the 5% level, * = significant at the 10% level.

Figure 3 shows a similar analysis based on different economic sectors following the classification of the Chilean Tax Authority. Sector 1 represents agricultural fishing and mining activities, Sector 2 industry and manufacturing, Sector 3 construction, Sector 4 retail activities, Sector 5 hotels and restaurants, Sector 6 regulated services (transportation, electricity,

sanitation, etc.), Sector 7 finance and business-related services, and Sector 8 all the remaining activities, including personal and social services. The largest estimated impacts are among sectors 4 and 8 with a 2.3% probability of paying higher taxes when receiving the notice. No statistically significant impact was found for Sectors 1, 3, 5 and 6.

Graph 3
Impact by Economic Sector



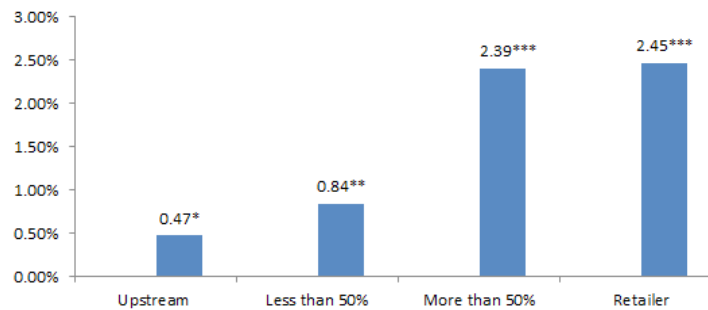
Notes: Each bar represents a separate regression for each economic sector. The numbers on top of the bars indicate the coefficient on being in the treatment group. The height of the bar represents the effect of receiving the notice as a percentage point increase in the probability of paying higher taxes than last year. *** = significant at the 1% level, ** = significant at the 5% level, * = significant at the 10% level.

The results are consistent with the idea that sectors with higher proportions of sales to final consumers (such as the retail sector), face higher incentives to underreport VAT, and therefore might react more strongly than sectors that have more transactions with other firms (such as agriculture, fishing and mining) or sectors that are highly regulated and thus face higher levels of scrutiny (such as sector 6).

Finally, Figure 4 shows the estimated reaction to the notice by firms with different shares of sales to final consumers as a fraction of total sales. We have grouped firms into 4

categories: 1) upstream: those that only sell to other firms; 2) firms that have some final sales but overall sell less than 50% of total sales to final consumers; 3) firms whose share of final sales are above 50% but below 100%; and 4) retailers: firms that sell only to final consumers. As Figure 4 shows, the estimated effect is stronger the larger the share of sales to final consumers. The difference in impact between retail firms and upstream firms is indeed substantial. While the response among retailers is a 2.45% increase in the probability of paying more taxes, the response is only 0.47% among upstream firms.

Graph 4
Impact by Share of Final Sales



Notes: Each bar represents a separate regression for each category of shares of sales to final consumers. The numbers on top of the bars indicate the coefficient on being in the treatment group. The height of the bar represents the effect of receiving the notice as a percentage point increase in the probability of paying higher taxes than last year. *** = significant at the 1% level, ** = significant at the 5% level, * = significant at the 10% level.

Similar to Pomeranz (2013), these results show that the effect of the notice is stronger on firms that have a larger proportion of sales to the final consumer. This is also in line with the self-enforcement hypothesis of the Value Added Tax. In the final stage of the production chain, the self-enforcing mechanism, which operates through the paper trail built into the VAT structure, breaks down: while it is in the firms' interest to ask suppliers for receipts in order to be able to deduct input costs from their VAT bill, final consumers have no incentive to do so (Keen, 2007).

2.3.2 Constructing a risk indicator

The analysis by subgroup shown above allows evaluating the extent of response for different groups of taxpayers. However, each taxpayer of course has many characteristics, and they are often correlated with one another. For example, retail firms also tend to be smaller firms. Tax authorities may therefore want to construct risk indicators that take into account many different characteristics of taxpayers, and predict for each taxpayer the likelihood of

responding to increased deterrence based on its characteristics. We present a very simple way to construct such a risk indicator based on regression analysis. We then briefly discuss how the results from the randomized deterrence messages can be used in combination with more sophisticated methods of data mining and prediction models.

To construct our simple risk indicator, we first calculate regression (2) for the entire sample with a set of taxpayer characteristics X_i . We then use the estimated coefficients $\hat{\beta}_1$ and $\hat{\beta}_3$ to calculate equation (3). $\hat{\beta}_3$ is a vector with a coefficient for each taxpayer characteristic, and will therefore be different for different firms. For an illustrative example, see the calculations in the appendix.

In our simple model, the value of the risk indicator ranges from -4.42% to 3.26%. The mean of the indicator is 1.04% and the standard deviation 1.02%. To get a sense of the differences in the magnitudes of the reaction to the notice among taxpayers, we can compare the value of the risk indicator at different parts of the distribution. The mean for the 99% percentile is 2.6 times larger than the mean for the entire sample, suggesting there are potentially large gains from targeting audits towards those firms that are more likely to increase tax payments when faced with a higher probability of being audited.

Many tax administrations are moving towards the use of more sophisticated prediction models, based on modern data mining techniques. The input from the randomized deterrence messages can be used as input into these models as well. To be able to combine the results from the randomized intervention with the data-mining

environment, an important additional step needs to be conducted.

The challenge lies in the fact that the prediction models require some risk measure to be associated directly with each observation. For example, tax administrations sometimes use information of whether a taxpayer has been detected for evasion in the past as one such measure. (As discussed above, these measures can frequently suffer from selection bias, because it is not random, which taxpayers were monitored to detect evasion in the past.) However, in the case the data obtained from the randomized experiment, there is no direct risk measure associated with each taxpayer. Instead, the information about how risky a taxpayer is only becomes apparent in comparison with the control group.

The following step allows bridging this gap: the prediction technology is applied twice, in a two-step process. First, tax payments for each firm are predicted, based on observable characteristics in the control group. Then, the difference is calculated for those in the treatment group, between the predicted value and the actual value. The predicted value is the value that would have been expected for the treated taxpayers, based on their observable characteristics, in the absence of treatment. The difference between actual and predicted can therefore be taken as a measure of the response to the deterrence message. This difference can now be used as input into the regular risk-based auditing models that the tax authority uses. For example, instead of (or in addition to) using a measure indicating whether a taxpayer has been detected for evasion in the past, the difference between actual and predicted can be used as a risk measure to build the risk indicator.

3. IMPLEMENTATION LESSONS LEARNED

In the course of implementing this methodology in Chile, a number of key lessons were learned. These lessons might also be useful for other tax authorities that are considering implementing a similar strategy, or for other organizations that conduct audits and can therefore follow the proposed methodology. The authors are also available to answer questions and provide support for practical implementation of this approach.

- **Selecting the right message:** The wording used in the deterrence letters not only needs to transmit effectively the increase in the probability of being audited, but should also avoid confusion among taxpayers and minimize the cases when taxpayers come to tax offices or call to request additional information. This allows minimizing both the cost of this intervention for the tax authority and the compliance costs for the taxpayers. Pretesting different types of wording and piloting the intervention can therefore be very useful to address these kinds of issues.
- **Ensuring the right taxpayers receive the letter:** To deal with this issue, the implementation of the methodology in Chile only included taxpayers with valid postal addresses. In addition, all the letters were sent via certified mail, which allowed for tracking those taxpayers who actually received the letter.
- **Trade-off between simplicity and prediction:** When estimating a reaction to the notice with regression analysis, it is important to determine what variables representing taxpayers' characteristics will be included in the regression analysis, and ultimately in the risk indicator. Since it is not clear ex-ante what kind of characteristics might have an effect on the outcome variable, one approach is to include in the regression as many potential explanatory variables as possible, and then rely on the statistical selection criteria to determine the right model.¹³ An alternative approach is to rely on a more simplified model that would include variables that are easy to interpret and analyze. In other words, there is a trade-off between increasing the predictive power of the model and relying on a more stable model in terms of time-consistency and resistance to noise in the data. More sophisticated data mining methods, as described above, are specifically designed to deal with such challenges, and to generate predictions based on large datasets.
- **Replications:** Some of the aforementioned lessons suggest that this methodology is better suited for certain contexts. For example, implementation would be more challenging in an environment where the tax authority lacks credibility, or where corruption limits the scope of audits to generate a real threat to taxpayers. On the other hand, implementation is easier where the tax authority has a strong reputation and better mechanisms are in place to credibly enforce audits.
- **Sample size vs. long-term credibility:** Being able to find statistically significant results requires a large number of treated firms. However, if taxpayers learn that a large number of notices were sent or that notices are sent repeatedly, they might assume that the tax authority has no intention or capacity to proceed with the auditing, or that the audits will not be very thorough. This could eventually attenuate or even eliminate any possible reaction from taxpayers (Ebrill et al 2001). In order to avoid such unintended

13. For a discussion on application of statistical selection criteria see Sala-I-Martin et al. (2004).

consequences, a tax authority that wishes to apply this methodology systematically and repeatedly should consider having the mailing followed by a number of actual detailed audits among taxpayers in the treatment group in order to produce a salient and credible increase in the likelihood of audit for these taxpayers. The tax authority might also consider sending a smaller number of letters regularly, rather than a very large number at the same time. The information acquired from these notices can then be aggregated over time in order to achieve the sufficient

sample size that provides the statistical power required for the risk indicator.

A combination of the mass-mailing strategy with a more long-term strategy can be to do a one-time mailing with a very large sample size in the first year, which gives baseline information for the risk indicator, and will work due to the one-time surprise. After that, the tax authority can use smaller mailings with a higher real increase in the audit probability to maintain credibility and at the same time assess possible changes in the responses over time.

4. COST-BENEFIT ANALYSIS

A simple back of the envelope calculation can show that constructing a risk indicator and using it for the purpose of optimizing auditing can be highly cost-effective. There are two types of benefits to the construction of a risk indicator: (i) a direct benefit of the increased revenue from the response to the notice, and (ii) subsequent benefits from using an optimized audit strategy based on the information provided by the risk indicator.

Similarly, there are also two types of costs: (i) a direct cost of distributing the deterrence notices and possible additional costs of enforcement measures associated with the deterrence notice, and (ii) a potential loss of credibility if messages that are not supported by real increases of the audit probability are sent out by the tax authority repeatedly (see discussion above on how to minimize this risk).

In terms of the direct returns associated with producing the indicator, we can compare the benefits of increasing tax revenues among the taxpayers that randomly received the deterrence notice to the actual cost of sending out the notice. When comparing total payments between treatment and control groups we found that the additional revenues for the first 6 months after

mailing the letters were in the order of 10 million USD, and after 18 months in the order of 16 million USD. On the other hand, the total costs of implementation, including the cost of sending out the letters, were in the order of 100,000 USD plus the costs related to implementing additional audits and administration. This is orders of magnitudes below the benefits obtained by the increase in tax revenues.

The subsequent benefits associated with the improved audit strategy are the result of targeting more heavily those firms with a higher estimated probability of paying higher taxes in response. As shown in the case of Chile, the large dispersion in the distribution of the risk indicator and how those differences are driven by specific firm characteristics provides scope to improve the targeting of auditing and reap benefits in terms of higher tax revenues. In addition, the tax authority was able to use the information obtained in the study for other purposes, such as gaining information about the duration of deterrence effects, improving the mailing system, and making budgetary calculations. Finally, many lessons learned and skills built during the implementation process have been employed fruitfully in subsequent internal studies and analyses.

5. CONCLUSIONS

This paper presents a new methodology to generate information that can be used to optimize the audit strategy of tax administrators. Based on a randomized evaluation design, this method allows estimating what types of taxpayers are more likely to react to an increase in the perceived probability of being audited. That information can then be used to construct a risk indicator. Tax authorities can use this risk indicator to optimize their audit strategies, which is often regarded as one of the main priorities for tax administration (Ebrill et al., 2001; Bird and Gendron, 2007).

This approach has two main advantages. First, it creates valuable information for tax administrators that is otherwise costly to obtain and usually only available for a small and non-representative set of taxpayers. This is especially relevant as the amount of information the tax authority has about different types of transactions in the economy has been found to be a key element that explains the difference in the performance of tax systems between developed and developing

countries (Gordon and Li, 2009). Second, unlike most risk indicators that make predictions based on past information and thereby are likely to suffer from self-selection and omitted-variable bias, our randomized design creates a reliable estimator of taxpayers' true responses to the audit threat.

The lessons from the application of our proposed methodology among VAT taxpayers in Chile show the importance of carefully designing and implementing the proposed analysis. Working with the appropriate sample size, ensuring the internal validity of the experiment, using the right wording to effectively change taxpayers' perceived probability of being audited, and truth-telling to maintain the credibility and deterrence power of the tax authority are all key elements to the success of the intervention. A similar approach can be used by tax authorities in many different contexts and for different taxes. Furthermore, it can easily be tailored to other types of organizations, whose operations require the implementation of audits or other monitoring processes.

6. BIBLIOGRAPHY

Adimark (2006). Imagen del Servicio de Impuestos Internos. Retrieved from http://www.sii.cl/aprenda_sobre_impuestos/estudios/imagen.htm

Bird, R. & Gendron, P. P. (2007). *The VAT in Developing and Transitional Countries*. New York: Cambridge University Press.

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *Handbook of Development Economics*, 4, 3895-3962.

Ebrill, L., Keen, M., Bodin, J. P. & Summers, V. (2001). *The Modern VAT*. International Monetary Fund, 1 (1), 1-23.

Keen, M. (2007). VAT Attacks!, *International Tax and Public Finance*, 14 (4), 365-381.

Keen, M. & Lockwood, B. (2010). The Value Added Tax: Its Causes and Consequences. *Journal of Development Economics*, 92 (2), 138-151.

Kleven, H. J., Kreiner, C. T. & Saez, E. (2009). Why Can Modern Governments Tax so Much? An Agency Model of Firms as Fiscal Intermediaries. NBER Working Paper, 15218.

Kopczuk, W. & Slemrod, J. (2006). Putting Firms into Optimal Tax Theory. American Economic Review: Papers and Proceedings, 96 (2), 130-134.

Pomeranz, D. (2013). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. NBER Working Paper 19199.

Sala-I-Martin, X., Doppelhofer, G. & Miller, R. I. (2004). Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach. The American Economic Review, 96 (4), 813-835.

Servicio de Impuestos Internos (2012). Ingresos Tributarios Anuales 1993 a 2012. Retrieved from <http://www.sii.cl/estadisticas/>

Vellutini, Ch. (2011). Risk-Based Audits: Approaches and Country Experience. The World Bank, Working Paper.

7. APPENDIX

Table A1
Results for our Illustrative Example of Regression (2)

	Coefficient	Robust Standard Errors
Treatment*Post	0.0052	0.0049
Metropolitan*Treatment*Post	0.0012	0.0022
Micro*Treatment*Post	0.0160***	0.0029
Small*Treatment*Post	0.0134***	0.0037
Medium*Treatment*Post	0.0025	0.0073
Sector1*Treatment*Post	-0.0226***	0.0053
Sector2*Treatment*Post	-0.0096*	0.0052
Sector3*Treatment*Post	-0.0105*	0.0057
Sector4*Treatment*Post	-0.0073*	0.0044
Sector5*Treatment*Post	-0.0194***	0.0063
Sector6*Treatment*Post	-0.0132**	0.0056
Sector7*Treatment*Post	-0.0095*	0.0053
Final Sales Share*Treatment*Post	0.0102***	0.0028
Exporter*Treatment*Post	-0.0195**	0.0082
Month Fixed Effects	Yes	
Number of Observations	7,308,631	
Number of Taxpayers	406,834	
R-squared	0.0004	

Notes: The table shows the results of linear probability regressions of an increase in declared VAT since the previous year. The first column shows the estimated coefficient and the second column the robust standard errors. *** = significant at the 1% level, ** = significant at the 5% level, * = significant at the 10% level.

Given the results shown in Table A1, the formula for the risk indicator is the following:

Risk Indicator = $0.0052 + 0.0012 \cdot \text{Metropolitan} + 0.0160 \cdot \text{Micro} + 0.0134 \cdot \text{Small} + 0.0025 \cdot \text{Medium} - 0.0226 \cdot \text{Sector1} - 0.0096 \cdot \text{Sector2} - 0.0105 \cdot \text{Sector3} - 0.0073 \cdot \text{Sector4} - 0.0194 \cdot \text{Sector5} - 0.0132 \cdot \text{Sector6} - 0.0095 \cdot \text{Sector7} + 0.0102 \cdot \text{Final Sales Share} - 0.0195 \cdot \text{Exporter}$

Variables Description

The variables used in the risk indicator and its categories are the following:

1. Dummy variable by regions: metropolitan, non-metropolitan
2. Dummy variables by size: no sales (omitted), micro, small, medium
3. Dummy variable by economic sector: sectors 1-8 as described in section 3.3.1 (sector 8 omitted)
4. Final sales share: continuous variables representing sales to final consumers as a proportion of total sales.
5. Dummy variables exporters: exporters, non-exporters (omitted)