

# Randomized Tax Enforcement Messages

## A Policy Tool for Improving Audit Strategies\*

Cristobal Marshall<sup>†</sup>

Dina Pomeranz<sup>‡</sup>

October 2013

### Abstract

Increasing tax revenue and reducing tax evasion is a key challenge for tax administrations around the world, particularly in developing countries. This paper presents a methodology to generate information to optimize audit strategies. Randomly selected taxpayers are presented with a deterrence message, and their subsequent tax payments are compared to a control group. This allows estimating what types of taxpayers are more likely to react to an increase in the perceived probability of being audited. Tax authorities can use this information to target audit activities towards categories of taxpayers that respond particularly strongly. They can also construct risk indicators that predict the likelihood of responding based on taxpayer characteristics. We present results from an application of this methodology among 445,000 firms in Chile, as well as the lessons learned during the implementation.

---

\*This project would not have been possible without the generous support of the following institutions: The Research and Studies Directorate at the Internal Revenue Service of Chile, 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. The authors are available for general or specific enquires on how to perform a similar evaluation in different contexts.

<sup>†</sup>Tierra Firme, Santiago, Chile; jmarshalls@gmail.com

<sup>‡</sup>Harvard Business School. Boston, USA; dpomeranz@hbs.edu.

# 1 Introduction

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). Planning an audit strategy that maximizes collection with a limited budget is therefore key for success, and the returns associated with such actions are potentially very high.

In this paper, we present a methodology for a critical component of an audit strategy: the selection of taxpayers to be audited. 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, parametric and non-parametric models are used to identify high-risk cases.<sup>1</sup> 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.

A key challenge of these methods, however, is 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, 2011) 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 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

---

<sup>1</sup>For a detailed description of risk-based methodologies see Vellutini (2011).

observable. Financial institutions have direct access to the information of credit default for all their clients, which, combined with clients' characteristics, can be used to develop a scoring model that predicts creditworthiness. Tax non-compliance, on the other hand, similarly 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.

Second, when data exists, for example for from prior audits, it is often difficult to gather conclusive insights from this data, because of selection bias. Audits are usually not randomly conducted. An analysis that compares data from audited with non-audited taxpayers faces such selection problems, since taxpayers in both 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 tax payer 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 where 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 do 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.

The issue of high costs of methods such as randomized audits is addressed, in that 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, increasing the incentives to declare more taxes. As tax payers may also vary in how they react to the threat of an audit, this strategy does not necessarily capture tax payers who are evading the most, but rather those respond most to an increase in the

probability of being audited. The results are thus aligned with a 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, 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 tax payer 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 of course be employed. The methodology consists of three steps:

1. *Deterrence notice*: The first and the 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 2 describes the methodology and the steps involved in its implementation. Section 3 provides background information and basic results from the intervention in Chile. Section 4 outlines some of the lessons learned during this implementation. Section 5 discusses the costs and benefits associated with the implementation of this methodology and Section 6 concludes.

## 2 Methodology

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

### 2.1 Implementation of the Deterrence Notice

The first step towards estimating taxpayers' response 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:<sup>2</sup>

- *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, importantly the universe needs to 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

---

<sup>2</sup>These steps usually also apply to the implementation of randomized evaluations in many other areas and contexts.

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 of backgrounds.

- *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 differences between different types of taxpayers, the sample size in both the treatment and 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*.<sup>3</sup> 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 will be. 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 does not necessarily have to 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 tax payers slated to receive the deterrence notice:* It is crucial for the validity of the analysis that taxpayers are selected in a random process among 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.<sup>4</sup> For this to remain true, it is important that subsequent to the random assignment, taxpayers in both treatment and control group continue to be treated exactly the same in all aspects except for the deterrence notice. Also, the control group has to consist only tax payers that were randomly selected no to receive the message. Any tax payers who were not part of the original universe, based on which the randomization was performed, cannot be included in the analysis.

---

<sup>3</sup>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).

<sup>4</sup>For more details on the rationale and benefits of using randomization see Duflo et al. (2007).

- *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 is preserved and not tampered with. This means two things in practice: first, insuring 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 is conducted based on this original random assignment.

## 2.2 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 and the control group.<sup>5</sup>

The overall impact of the message can be analyzed in regression analysis, following the specification shown in Equation (1). This specification is a difference-in-difference

---

<sup>5</sup>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 their actual treatment status (intention to treatment, 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.

approach. A simple difference of the post-treatment outcomes would also be valid. The inclusion of the pre-treatment period and of 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  $t$ ,  $Z_i$  is a dummy variable that indicates whether the taxpayer has been assigned to the treatment group,  $tZ_i$  is a dummy that indicates whether the taxpayer belong to the treatment group in the post-treatment period,  $\delta_t$  stands for a year or month fixed effect,  $\lambda_i$  is a taxpayer fixed effect and  $e_{it}$  is an error term. This regression estimates the coefficient of interest,  $\hat{\beta}_1$ , which represents the overall treatment effect, i.e. the difference in the outcome of interest between taxpayers in the treatment and control groups.

Even when interested in the effect of the notice on one particular outcome, such as declared income tax, different independent variables can be chosen for estimation purpose, including:

- *The total amount of taxes declared:* Provides an easy interpretation of the estimated effects, but the magnitudes of the estimations are driven by the relative size of taxpayers, making it less useful when interested in estimating differential effects by type of taxpayers (i.e. small vs. medium size firms).
- *The log of the amount of taxes declared:* Could be used if the data is non-normally distributed, but is less useful when the sample includes observation with zeros, or non-declaring taxes.
- *The median of declared taxes:* A quantile regression can be used when there is large dispersion in the data, since the median is less sensitive to extreme values than the mean.
- *Probability of declaring positive taxes:* If the margin of interest is whether firms are declaring taxes after receiving the notice instead of changes in the level of declared taxes, a linear probability regression with the probability of paying positive taxes as the independent variable can be used.
- *Probability of declaring higher taxes compared to last year:* Similarly, a linear probability regression with the probability of declaring higher taxes than last year as

independent variable might be desirable when, in addition to the argument made above, there is a fat tail in the distribution of declared taxes.

The choice of the variable to be used for estimation purpose should be made on a case-by-case basis depending on the data available 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 we interpret the response to the letter? Can the magnitude of the taxpayers' reaction 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 tax 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, 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 probability of being audited.

## **2.3 Extracting information for improved audit targeting**

### **2.3.1 What types of taxpayers respond more?**

This approach not only allows us to estimate the magnitude of taxpayers' overall reaction 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, and 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 taxpayer's 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 actually reveal evasion, and (v) the taxpayer's degree of risk aversion, to name a few. A higher response therefore does not necessarily imply a higher level of

evasion, but a higher response to an increase in the audit probability.

For the estimation of the differential response by type of taxpayer, we use a variation of regression ??, 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's characteristics, and  $tZ_i X_i$  is an interaction term of firm 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. For example, we can classify taxpayers by level of sales according to four categories: no sales, micro size, small size and medium size. Then,  $X_m$ ,  $X_s$ ,  $X_M$  represent dummy variables of whether the taxpayers are classified as micro, small or medium size taxpayer respectively (no sales is the omitted category). Then, the estimated coefficient  $\hat{\beta}_3$  for the  $tZ_i X_m$  variable represents the differential impact of the deterrence letter in micro taxpayers compared to the effect on taxpayers that show no sales. A positive estimated coefficient means, everything else equal, micro size taxpayers 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.

### 2.3.2 Constructing a risk indicator

Rather than analyzing the differential effects for specific taxpayer characteristics in isolation, we can also aggregate the information how taxpayer characteristics interact with the treatment into one single risk indicator. This can be done using the estimated coefficients for  $\beta_1$  and the differential effects by taxpayers' types  $\beta_3$  from regression 2. Together, these predict the response to the deterrence notice by each taxpayer based on his 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 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 last 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.

### 3 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 provision of high-quality public service, constant service improvement, and application of cutting-edge methods and technologies. Following that 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 in practice in Chile.

#### 3.1 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, 2010). Although VAT is believed to have distinctive features that make it less vulnerable to evasion and fraud, in recent years there has been 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 analysis consisted of all small and medium size firms in Chile that were operating in June 2008, declared a positive amount of VAT for at least one month between July 2007 and June 2008, and had valid postal addresses. Larger taxpayers were not included in the analysis because they face a different audit process administered by a separate unit within SII.

- *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.<sup>6</sup> In particular, the letters informed taxpayers that in a random selection process among micro, small and medium size firms, they have 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 taxpayers, asking them 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, 102,000 were required to receive the deterrence message. As argued in section 2, treatment and control groups do not necessarily have to have the same number of units.
- *Random selection of those tax payers 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

---

<sup>6</sup>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 (2011).

groups comparable in all characteristics, the sample was stratified by taxpayer size. Randomization then took place within these subgroups.

- *Pilot*: A pilot of 250 letters was sent out 2 months prior to the large-scale mailing, helping detect wrinkles in the process. It allowed adjusting of some internal procedures and wording of the letters, and assured a smooth implementation of the large scale mailing.
- *Ensuring that the random selection is respected in the implementation*: 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 treated group received no letter, taxpayers with invalid postal address were excluded from the universe. This allowed for a substantial increase in the statistical power of the experiment, by significantly reducing non-compliance with treatment assignment.<sup>7</sup>

## 3.2 Estimating the reaction to the notice

### 3.2.1 Summary statistics at baseline

The first step of 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 previous to the intervention. Table I 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.

First, we look at whether taxpayers are balanced in terms of tax payments previous to sending the notice. The group of firms that take 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 empirical analysis.

---

<sup>7</sup>Partial compliance strongly reduces the power of the evaluation design. The higher the non-compliance rate, the higher the sample size that is required to estimate a minimum effect.

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 from the previous tax year: firms with no sales, micro-sized, small-sized, medium-sized and large-sized firms.<sup>8</sup> The micro-sized taxpayers are by far the largest group, comprising 74.5% of the universe. The second largest groups are 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.

### 3.2.2 Summary statistics at baseline

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 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 measure has several benefits compared to alternative measures. First, as opposed to log specification, it can be applied to variables that include zero or negative values, which is the case for declared taxes. Second, it is robust outliers, as is the case with our fat tail 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 which are sensible to taxpayers size), making it useful to compare the treatment effect between different types of firms.<sup>9</sup>

---

<sup>8</sup>Large firms are not included in this study.

<sup>9</sup>See Pomeranz (2011) for a detailed discussion on the use of alternatives dependent variables, and why this particular outcome variable was chosen in the Chilean case.

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, 2011).

Two other interesting results from the evaluation, as shown in Pomeranz (2011), are worth noticing. First, the impact in terms of higher taxes seems to be very persistent. The initial impact in the treatment group is immediate and slowly decreases over the following months, reaching the same level as the control group after about 18 months. In the first six months of the implementation of the enforcement letter, the increase in tax revenue is estimated in 3.3 million USD. Second, increasing the audit probability of firms generates 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 firms trading network.

### **3.3 Extracting information for improved audit targeting**

#### **3.3.1 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 I shows that the estimated effect of the deterrence letter is higher in the Metropolitan Region and the South and lower in the North. 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 II 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., 2010). Internal processes that keep track of transactions, monitoring mechanisms, and reputational effects are more likely to be present among larger taxpayers, making them less likely to respond to an external 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 significant effect among medium size taxpayers.

Figure III 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.

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 than sectors that are highly regulated, and thus face higher levels of scrutiny (such as sector 6).

Finally, Figure IV shows the estimated reaction to the notice by firms with different shares of final sales as a fraction of total sales. We have grouped firms as upstream or those that only sell to other firms, firms that sell less than 50% of total sales to final consumers, firms whose share of final sales are above 50% and retailers who only sell to final consumers. The estimated effect is stronger the higher the share of final shares, with a very large differences in impact between retail taxpayers upstream taxpayers. 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 taxpayers.

Similar to Pomeranz (2011), these results show that the effect of the notice is stronger on firms that have a larger proportion of sales to the final consumer, in line with the self-enforcement hypothesis of the Value Added Tax. In this final stage of the production chain, the self-enforcing mechanism, which is carried 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).

### 3.3.2 Constructing a risk indicator

To construct our risk indicator we calculate regression (3) for the entire sample using the estimated coefficients  $\hat{\beta}_1$  and  $\hat{\beta}_3$  and a set of variables representing taxpayers' characteristics.

The values of the risk indicator range 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 on targeting audits towards those firms that are more likely to increase tax payments when faced with a higher probability of being audited.

## 4 Implementation lessons learned

In the course of implementing of this methodology in Chile, a number of key lessons were learned. These lessons might also be useful for other tax authorities considering implementing a similar strategy or for other organizations that also conduct audits and can therefore follow the proposed methodology.

- *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 it also has to avoid confusion among taxpayers and minimize the cases when taxpayers come to tax offices or call to request additional information. Pretesting different types of wording and piloting the intervention can both be very useful when addressing these kinds of issues.
- *Assuring 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, allowing for tracking those taxpayers who actually received the letter.
- *Trade-off between simplicity and prediction:* When estimating a reaction to the notice 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.<sup>10</sup> 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.

- *Interpretation of the response:* As argued above, an increase in tax payments as a response to the deterrence notice indicates a larger propensity to increase compliance in response to an increased audit probability but not necessarily a higher level of evasion. While firms with stronger responses are more likely to have higher levels of evasion, there are other aspects discussed in section 2 that also determine the response level. When relying on this methodology to target audit strategy, tax administrators need to be aware of other strategies that may be required in order to find those high evasion taxpayers with low risk indicators.
- *Replications:* Some of the aforementioned lessons suggest that this methodology can be better implemented in some contexts. For example, implementation would be more challenging in an environment where tax declaration and payments are difficult to enforce, where the tax authority lacks credibility, and where corruption limits the scope of audits to generate a real threat to taxpayers. On the other hand, implementation would be easier when the tax authority has a strong reputation and better mechanisms in place to enforce an announced increase in audits.
- *Sample size vs. long-term credibility:* Being able to find statistically significant results requires the right sample size. However, if taxpayers learn that a large number of notices were sent or if notices are sent repeatedly, they might assume that the tax authority actually has no intention to proceed with the auditing, or that the audits will not be sufficiently thorough. This could eventually attenuate or even eliminate any possible reaction from taxpayers (Ebrill et al 2001). In order to avoid such cases, a tax authority that wishes to apply this methodology systematically and repeatedly could pursue the following strategy: send a smaller number of notices each year followed by actual detailed audits, to produce a salient and credible

---

<sup>10</sup>For a discussion and application of statistical selection criteria see Sala-I-Martin et al. (2004)

increase in the likelihood of audit for these taxpayers. The information acquired from these audits can then be aggregated over time in order to achieve the sufficient sample size that provides the statistical power required for the risk indicator. An appropriate combination of the mass-mailing and the more long-term strategy might 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.

## 5 Cost - Benefit Analysis of the Methodology

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 costs of distributing the deterrence notices and possible additional costs of enforcement measures associated with the deterrence notice, and (ii) a possible loss of credibility if frequently messages are sent out by the tax authority that are not supported by real increases of the audit probability (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 includes 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, which in sum is order 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. 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 a scope to improve the targeting of auditing and reap benefits in terms of higher tax revenues. The costs associated to the loss of credibility are hard to assess and are not part of the aim of this paper.

## 6 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 us to estimate 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 then use this risk indicator to optimize their audit strategies, which is often regarded as one of main priorities for tax administration (Ebrill et al., 2001; Bird and Gendron, 2007).

Our 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 response 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 for the success of the intervention. A similar approach can be used by other tax authorities in many different contexts and for different

taxes. Furthermore, it can easily be tailored other types of organizations whose operations require the implementation of audits and other monitoring processes.

## References

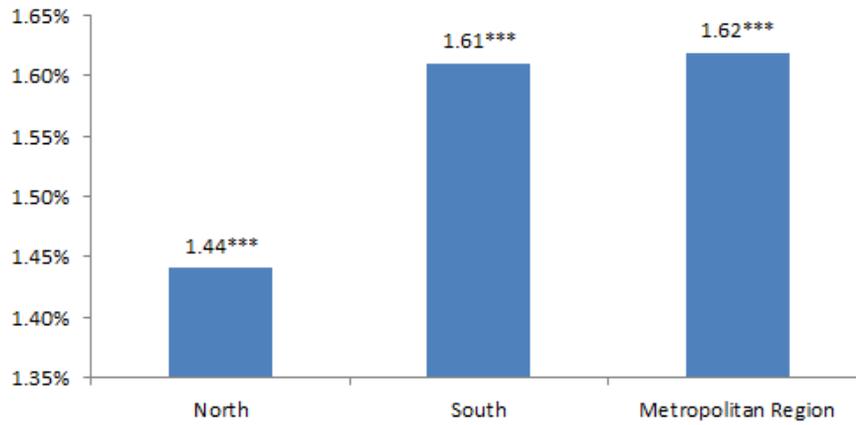
- Adimark**, “Imagen del Servicio de Impuestos Internos,” *Journal of Public Economics*, 2006.
- Bird, Richard and Pierre-Pascal Gendron**, “The VAT in Developing and Transitional Countries,” *Cambridge University Press*, 2007.
- Duflo, Esther, Rachel Glennerster, and Michael Kremer**, “Using Randomization in Development Economics Research: A Toolkit,” *Handbook of development economics*, 2007, 4, 3895–3962.
- Ebrill, Liam, Michael Keen, Jean-Paul Bodin, and Victoria Summers**, “The Modern VAT,” *International Monetary Fund*, 2001, 1 (1), 1–23.
- Gordon, Roger H. and Wei Li**, “Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation,” *Journal of Public Economics*, 2009, 93 (7-8), 855–866.
- Keen, Michael**, *VAT Attacks!*, Vol. 14, Springer, August 2007.
- **and Ben Lockwood**, “The Value Added tax: Its Causes and Consequences,” *Journal of Development Economics*, July 2010, 92 (2), 138–151.
- Kleven, Henrik J., Martin Knudsen, Claus T. Kreiner, Soren L. Pedersen, and Emmanuel Saez**, “Un-willing or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark,” *Econometrica*, 2010, 79 (3), 651–692.
- Kopczuk, Wojciech and Joel Slemrod**, “Putting Firms into Optimal Tax Theory,” *Papers and Proceedings*, 2006, 96 (2), 130–134.
- Pomeranz, Dina**, “No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax,” 2011. Working Paper.
- Sala-I-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller**, “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach,” *The American Economic Review*, 2004, 96 (4), 813–835.
- Servicio de Impuestos Internos**, “Ingresos Tributarios Anuales 1993 a 2009,” 2010. Working Paper.
- Vellutini, Charles**, “Risk-Based Audits: Assessing the Risk,” *Chapter 2 in Risk-Based Audits: Approaches and Country Experience*, *The World Bank*, 2011. Working Paper.

## 7 Figures and Tables

**Table 1:** Baseline Summary Statistics and Balance of Randomization

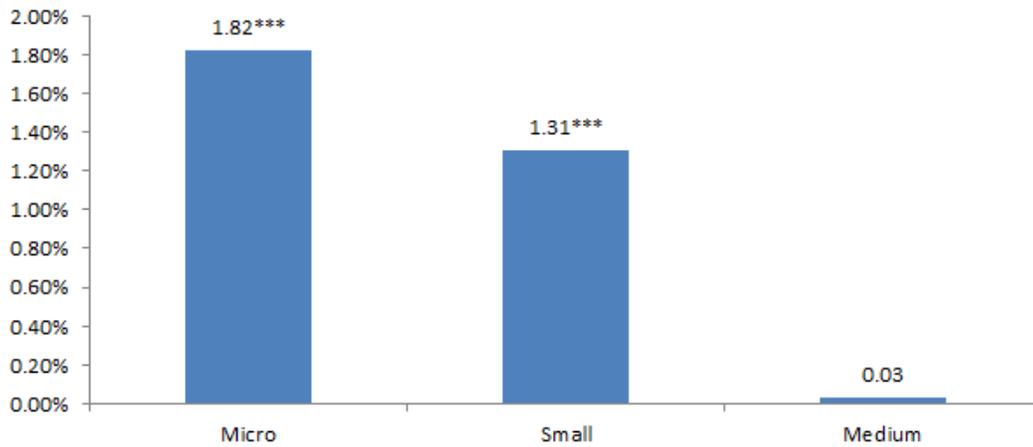
|                           | Control<br>Group   | Difference<br>Treatment Group |
|---------------------------|--------------------|-------------------------------|
| Monthly tax paid (mean)   | 264,434<br>(2,746) | 1,342<br>(10,144)             |
| Monthly tax paid (median) | 69,892<br>(458)    | -779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

*Notes:* Each row shows the result obtained from regressing the variable in question on the treatment dummy variables and the constant term. The constant term captures the value that corresponds to the control group. Robust standard errors (in parentheses) are clustered at the taxpayer level. Monetary amounts are in Chilean pesos (where 500 pesos are approximately equivalent to 1 USD). None of the differences are statistically significant at the 10% significance level. All the data (including the taxpayers' characteristics) was obtained from official tax records.



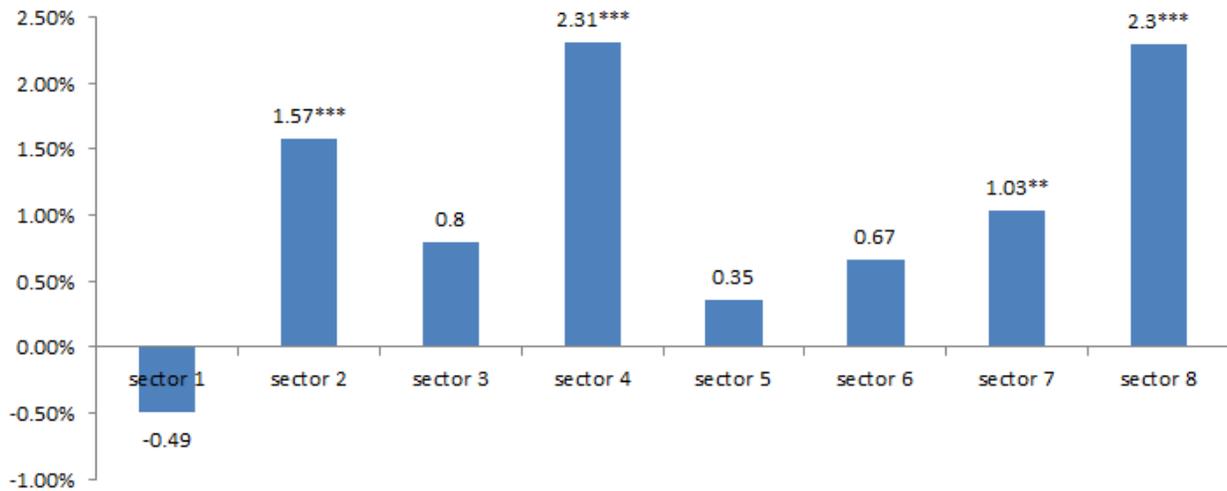
**Figure 1:** Impact of the notice by region

*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.



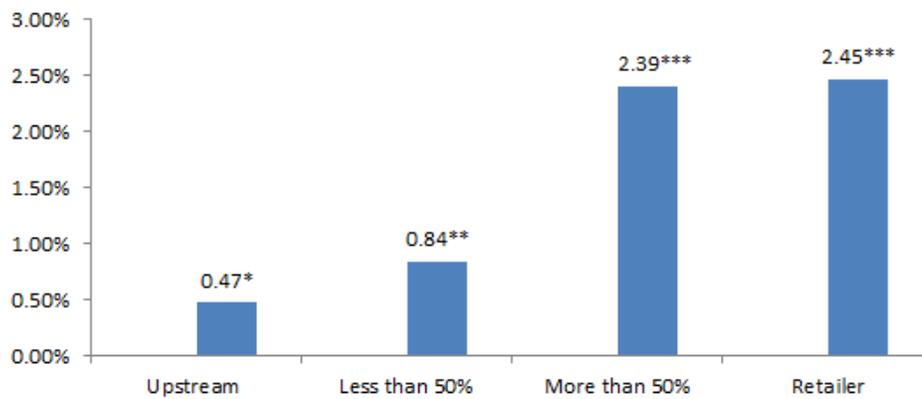
**Figure 2:** Impact of the notice by firm size

*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:** Impact by Economic Sector

*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 4:** Impact by Share of Final Sales

*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.

## APPENDIX

### A Estimations

**Table A1:** show the results for our preferred specification for regression 2

|                                  | Coefficient | Robust<br>Standard Errors | t-statistic |
|----------------------------------|-------------|---------------------------|-------------|
| Treatment*Post                   | 0.0052      | 0.0049                    | 1.06        |
| Metropolitan*Treatment*Post      | 0.0012      | 0.0022                    | 0.57        |
| Micro*Treatment*Post             | 0.0160***   | 0.0029                    | 5.61        |
| Small*Treatment*Post             | 0.0134***   | 0.0037                    | 3.67        |
| Medium*Treatment*Post            | 0.0025      | 0.0073                    | 0.35        |
| Sector1*Treatment*Post           | -0.0226***  | 0.0053                    | -4.28       |
| Sector2*Treatment*Post           | -0.0096*    | 0.0052                    | -1.83       |
| Sector3*Treatment*Post           | -0.0105*    | 0.0057                    | -1.84       |
| Sector4*Treatment*Post           | -0.0073*    | 0.0044                    | -1.65       |
| Sector5*Treatment*Post           | -0.0194***  | 0.0063                    | -3.09       |
| Sector6*Treatment*Post           | -0.0132**   | 0.0056                    | -2.35       |
| Sector7*Treatment*Post           | -0.0095*    | 0.0053                    | -1.8        |
| Final Sales Share*Treatment*Post | 0.0102***   | 0.0028                    | 3.64        |
| Exporter*Treatment*Post          | -0.0195**   | 0.0082                    | -2.37       |
| 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 showed in Table A.1, the formula for the risk indicator is the following:

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

## B 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 Variable by Size:

- No sales (omitted)
- Micro
- Small
- Medium

3. Dummy Variable by Economic Sector:

- Sector 1: Agricultura, Ganazadera, Caza, Silvicultura, Pesca y Minera
- Sector 2: Industrias Manufactureras
- Sector 3: Construccin
- Sector 4: Comercio
- Sector 5: Hoteles y Restaurantes
- Sector 6 : Transporte, Comunicaciones, Electricidad, Gas y Agua
- Sector 7: Intermediacin financiera y servicios empresariales
- Sector 8: Servicios sociales, personales y otros (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)