Re-Assessing the Merits of Measuring Tax Evasions through Surveys: Evidence from Serbian Firms

Thorben C. Kundt, Florian Misch, and Birger Nerré

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Re-Assessing the Merits of Measuring Tax Evasions through Surveys: Evidence from Serbian Firms

Thorben C. Kundt, Florian Misch,
and Birger Nerré

Non-Technical Summary

While there is no doubt that tax evasion is widespread in developing countries, exact measures of its extent are difficult to obtain because, by definition, tax evasion is hidden. However, such estimates are often critical from a policy perspective, for instance to assess the pay-offs of costly countervailing measures. One obvious and potentially timely source of information are taxpayer surveys. However, data on self-reported tax evasion obtained through surveys are typically not reliable. The reason is that survey respondents can be expected to answer questions about tax evasion (or other sensitive topics) not truthfully out of fear of the negative consequences that this may entail if their answers are revealed.

To overcome this inherent problem of surveys, sophisticated questioning methods have been developed which protect the privacy of the individual responses by 'bundling' sensitive and non-sensitive questions. In other words, these methods do not allow the interviewer to identify whether, and to what extent, a particular respondent shows the behavior in question. This is likely to provide incentives to the respondent to answer truthfully even if the question is sensitive. At the same time, these questioning methods still allow estimating the prevalence of the sensitive behavior across the whole sample of respondents. These methods have been mostly applied by researchers in sociology and psychology to study topics such as drug use and sexual behavior.

The objective of this paper is to evaluate whether a novel questioning method of this type referred to as the crosswise model allows obtaining more reliable estimates about tax evasion through business surveys. The main advantage of the crosswise model is that, compared to similar methods, respondents are not able to resort to an obvious self-protective strategy other than refusal to answer. We compare this method to the conventional approach taken in other business surveys such as the World Bank Enterprise Surveys using data from a rich firm survey in Serbia carried out in November and December 2012. Contrary to most other surveys that examine tax evasion, we also differentiate between two types of tax evasion, namely underreporting of sales and of wages (envelope wages).

We randomly split our sample into two subsamples which are almost identical in terms of their industry-size-region distribution. For both subsamples, we estimate the extent of sales and wage underreporting, in one case using the CM, and in the other case using the 'conventional' approach. Under the latter, the framing of the question is 'forgiving', and respondents are asked to refer to firms that are similar to theirs as an imperfect but standard way to entice truthful answers. In line with our predictions, we show that the estimated shares of firms that underreport sales and wages by at least 10% are both higher under the CM, although the difference is only robustly significant for the underreporting of sales. We conclude that the CM – through fully protecting the privacy of respondents – provides better information about tax evasion and should be increasingly used in business surveys.

Das Wichtigste in Kürze

Es besteht kein Zweifel, dass in Entwicklungsländern Steuerhinterziehung weitverbreitet ist. Dennoch ist es schwierig, ihren genauen Umfang zu schätzen, da sie definitionsgemäß im Verborgenen stattfindet. Um den Nutzen von Gegenmaßnahmen beurteilen zu können, sind jedoch solche Schätzungen unabdingbar. Eine naheliegende Informationsquelle, die zudem zeitnah Schätzungen liefern kann, sind Befragungen von Steuerzahlern. Allerdings sind Befunde, die auf der Selbsteinschätzung der Befragten basieren, naturgemäß nicht verlässlich, da die Befragten aus Angst vor potentiellen negativen Konsequenzen Fragen zum eigenen Steuerverhalten möglicherweise nicht wahrheitsgemäß beantworten.

Um dieses generelle Problem von Befragungen zu umgehen, wurde eine Reihe elaborierter Fragemethoden entwickelt, die darauf abzielen, durch „Bündelung" sensibler und nicht-sensibler Fragen die Privatsphäre der Befragten zu schützen. Da der Interviewer keine Aussagen darüber treffen kann, ob und in welchem Umfang ein bestimmter Befragter das spezifische Verhalten zeigt, ergibt sich für diesen ein Anreiz, selbst sensible Fragen wahrheitsgemäß zu beantworten. Gleichzeitig erlaubt diese Methode eine Schätzung der Prävalenz des Verhaltens in der gesamten Stichprobe. Diese Fragetechniken wurden bisher hauptsächlich in psychologischen und soziologischen Studien zur Untersuchen von Themenfeldern wie Drogenmissbrauch oder Sexualverhalten angewandt.

Das Ziel dieser Studie besteht darin zu bewerten, inwieweit Umfragen, die auf dieser sogenannte Crosswise-Methode (CM) basieren, verlässlichere Informationen zur Steuerhinterziehung als herkömmliche Umfragemethoden liefern. Der zentrale Vorteil der CM besteht darin, dass die Befragten im Gegensatz zu ähnlichen Umfragemethoden - abgesehen von Antwortverweigerung - auf keine andere Strategie zum Selbstschutz zurückgreifen können. Wir vergleichen die CM mit üblichen Verfahren, die in anderen Unternehmensumfragen wie den World Bank Enterprise Surveys angewendet werden, auf Basis einer Befragung serbischer Unternehmen, die Ende 2012 durchgeführt wurde. Anders als die meisten Unternehmensbefragungen differenzieren wir zwischen zwei verschieden Arten der Steuerhinterziehung.

Wir unterteilen unsere Stichprobe durch zufällige Ziehung in zwei Teilstichproben, wobei deren Struktur hinsichtlich der Unternehmensbranchen, Unternehmensgrößen und Region der Unternehmensstandorte in beiden Teilstichproben sehr ähnlich ist. In beiden Stichproben wird das Ausmaß der Steuerhinterziehung geschätzt, in einem Fall mit Hilfe der CM, in dem anderen Fall mit Hilfe eines konventionellen Ansatzes. Im letzteren Fall wird in der Formulierung der Frage Steuerhinterziehung als ‚nicht schlimm' dargestellt, und die Befragten werden gebeten, sich in ihrer Antwort auf vergleichbare Unternehmen zu beziehen. Die vorliegende Studie zeigt, dass der geschätzte Anteil serbischer Unternehmen, die mindestens 10% ihrer Umsätze bzw. Lohnsummen gegenüber der Finanzverwaltung verschweigen, im Falle der CM höher ausfällt. Der Unterschied ist allerdings nur hinsichtlich der Umsätze signifikant. Die Studie schlussfolgert, dass die CM eine realistischere Schätzung von Steuerhinterziehung erlaubt, da die Privatsphäre der Befragten besser geschützt wird.

Re-Assessing the Merits of Measuring Tax Evasions through Surveys: Evidence from Serbian Firms[*]

Thorben C. Kundt[a] Florian Misch[b] Birger Nerré[c]

July 2013

Abstract

This paper addresses the major weakness of measuring tax evasion through business and household surveys, namely the reluctance of respondents to answer truthfully due to the threat of disclosure. First, we assess the merits of a novel questioning method to gather information about tax evasion by means of business surveys. This approach allows estimating the prevalence of tax evasion, but it does not allow identifying whether the individual firm engages in tax evasion or not, therefore providing incentives for survey participants to answer truthfully. Second and contrary to most other business surveys, we differentiate between two common modes of tax evasion, namely underreporting of sales and informal supplements to official wages ('envelope wages'). Using evidence from Serbia, we show that the estimated share of firms which underreport sales and wages, respectively, by at least 10% is higher under the crosswise model compared to the case when conventional questioning methods applied in business surveys such as the World Bank Enterprise Surveys are used. However, the difference is only significant with respect to sales. These results appear to be robust to a number of modifications, and we explore various potential causes that lead to these results.

JEL codes: H20, E62

Keywords: tax evasion, shadow economy, measurement, developing countries

[a] Address: Helmut-Schmidt-University, Hamburg, Germany. E-mail: kundt@hsu-hh.de

[b] Corresponding author. Address: Centre for European Economic Research (ZEW), Mannheim, Germany. E-mail: misch@zew.de

[c] Address : Public Finance Reform, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, Belgrade, Serbia. E-mail: birger.nerre@giz.de

1   Introduction

While there is no doubt that tax evasion is wide-spread in developing countries, exact measures of its extent are difficult to obtain because, by definition, tax evasion is hidden. However, such estimates are often critical from a policy perspective, for instance to assess the pay-offs of costly countervailing measures. One obvious and potentially timely source of information are surveys, but it has long been recognized that respondents in such surveys have strong incentives to not answer truthfully questions about tax compliance due to the threat of disclosure.[1] The objective of this paper is to assess whether this weakness of surveys can be overcome by means of a novel questioning method referred to as the crosswise model.

Of course, there exist at least two other approaches at the microeconomic level to assess the extent of tax evasion which are summarized by Gemmell and Hasseldine (2012) and Alm (2012) and which are part of a broader trend in economics to find evidence of hidden or illicit behavior in a variety of settings; see Zitzewitz (2012) for a general survey.[2] While these approaches are compelling from a conceptual perspective, they may not always be applied in a developing countries' context. First, intensive taxpayer audits may be the most obvious strategy to gain information about the extent of tax evasion, but they are subject to various caveats: they are costly and, as Slemrod and Weber (2012) suggest, there is no guarantee that such tax audits detect all unreported income, especially in an environment with a high presence of non-tax filers, strong reliance on cash in business transactions and dishonest or insufficiently monitored tax inspectors.

The second strategy is to use data on declared income of wage earners and self-employed and to exploit the assumption that both groups show identical behavior, except that wage earners report income truthfully contrary to the self-employed who do not, given that they have more opportunities to evade taxes. However, the assumption of this strategy, namely that wage earners report income truthfully contrary to the self-employed who do not, is likely to be violated in developing countries where firms are likely to evade payroll tax and social security contributions by underdeclaring wages as well. For instance, Feldman and Slemrod (2007) compare the correlations between charitable contributions and reported income for both groups. In order to study tax evasion of the self-employed, they assume that differences in this relationship

---

[1] See for instance Long and Swingen (1991). In his paper, Slemrod (2007, p.25) asks the reader provocatively: "Would you answer survey questions about tax evasion honestly?" Hessing et al. (1988) compare self-reported non-compliance with findings of intensive audits and find no correlation between both which is likely to result from dishonest survey responses.

[2] This branch of the literature is somewhat distinct from the literature examining the determinants of tax evasion; see Torgler (2011) for a survey.

across income types are solely due to underreporting; see Pissarides and Weber (1989) for a similar approach.[3][4]

In this paper, we assess tax evasion by simply asking firm managers in a business survey if they evade taxes. We employ a particular questioning strategy to entice truthful answers which is part of a broader class of methods that dates back to Warner (1965) and that is summarized by Tourangeau and Yan (2007) to study sensitive topics in sociology and psychology such as drug use and sexual behavior, where researchers face similar problems. Generally speaking, these methods protect the privacy of the individual responses by 'bundling' sensitive and non-sensitive questions so that the interviewer does not know whether a particular respondent engages in the type of illicit or sensitive behavior. This, in turn, is likely to provide incentives to the respondent to answer truthfully even if the question is sensitive. At the same time, these methods still allow estimating the prevalence of the sensitive behavior across the whole sample of respondents.

The first contribution of this paper is a methodological one: we assess the merits of one particular and novel method of this class, namely the crosswise model (CM) originally proposed by Yu et al. (2008) and applied to plagiarism by Jann et al. (2012), to study tax evasion in business surveys. We review the mechanics of the CM and its advantages in the next section. While this class of methods has been used to study a variety of topics in the social sciences, they have been rarely applied to tax evasion, and there are neither studies that focus on developing countries nor studies that make use of the CM - which we argue is by far best suited in this context. Our benchmark method against which we evaluate the CM is the conventional approach to study sensitive issues including tax evasion used in the World Bank Enterprise Surveys (WBES), which is the most widely used source of standardized firm-level data from developing countries. In the WBES, the techniques used only imperfectly protect the privacy of the participants' answers. In most cases, the questions are solely framed in a way that justified tax evasion based on economic grounds, and/or respondents are asked to refer to firms / households that are similar to themselves. However, respondents are still likely to realize the rationale and may answer dishonestly or do not answer at all. This implies that these surveys are likely to underestimate the extent of illicit behavior including the extent of tax evasion.

---

[3] A related strand of the literature studies the effects of enforcement and audit threats on tax compliance behavior. Almunia and Lopez-Rodriguez (2012) exploit a discontinuity in enforcement in Spain that arises as taxes from firms above a certain size threshold are collected by a large taxpayer unit which also monitors firms more intensively. By contrast, Slemrod et al. (2001) studies the effects of threat letters send randomly to taxpayers.

[4] There are also macroeconomic approaches often referred to as indirect strategies which are used to estimate the size of the overall shadow economy which is broader than the concept of tax evasion; see Schneider and Enste (2000) for a survey of the literature. Henderson et al. (2012) is one of the more innovative papers taking a macro approach: they use luminosity as measured from outer space as an indicator of 'true' economic activity.

Our second contribution is that, contrary to most other surveys of firms and households, especially those that employ related methods, we differentiate between two types of tax evasion that are both common in developing countries, namely underreporting of sales to evade sales and/or profit taxes, and 'envelope wages' to evade payroll taxes and/or social security contributions.[5] Sales underreporting often implies that the seller does not issue invoices and typically asks for cash payment. Envelope wages imply that employers top up the official wages using undeclared cash payments; the latter type of tax evasion is often referred to as 'envelope wages'. From a policy perspective and to design strategies to increase tax compliance, distinguishing the modes of tax evasion is important.

Our data come from a recent survey of small and medium Serbian firms. In Serbia, the shadow economy is relatively large compared to other countries in the region according to Krstić et al. (2013) and envelope wages are common, given the high proportion of workers that are paid exactly the minimum wage.[6] We randomly split the sample into two subsamples which are almost identical in terms of their industry-size-region distribution. For both subsamples, we estimate the extent of sales and wage underreporting, in one case using the CM, and in the other case using the 'conventional' approach. Under the latter, the framing of the question is 'forgiving' and respondents are asked to refer to firms that are similar to theirs as an imperfect but standard way to entice truthful answers. In line with our predictions, we show that the estimated shares of firms that underreport sales and wages by at least 10% are both higher under the CM, although the difference is only robustly significant for the underreporting of sales. We conclude that the CM – through fully protecting the privacy of respondents – provides better information about tax evasion and should be increasingly used in business surveys.

The paper is organized as follows. Section 2 provides an overview of survey techniques for sensitive topics and its applications to tax evasion, and it describes the crosswise model and its advantages in detail. Section 3 describes the data and our survey design. Section 4 presents the empirical specification and results. Section 5 concludes.

---

[5] Several empirical studies on the determinants of tax evasion and tax morale rely on data from the World Values Survey (WVS) (see for example Torgler, 2011). The WVS only contains one relevant question about tax evasion, namely respondents are asked to evaluate how "justifiable" tax evasion in general is.

[6] Employers and employees may benefit from envelope wages and therefore jointly agree on this practice; see Schmidt and Vaughan-Whitehead (2011) for a discussion of envelope wages in Serbia and other countries of the region.

2 Literature Review

This section reviews various methods to assess illicit or otherwise sensitive behavior in surveys in general and some of the survey-based evidence on the extent of tax evasion. Under the most obvious method, referred to as 'direct questioning' (DQ), respondents are directly asked whether they engage in the sensitive behavior in question. However, at least since Barton (1958), several more sophisticated survey techniques have been developed that aim at eliciting truthful answers to questions about sensitive topics (see Table 1 for an overview). These can loosely be divided in more elaborated approaches that protect the privacy of the individual answers and conventional approaches which do not. We briefly review these techniques before discussing in detail the crosswise model which we argue is the most suitable to estimate the prevalence of tax evasion; see Tourangeau and Yan (2007) for a more detailed summary of these techniques.

2.1 Conventional Approaches

In economics, the standard methods include the use of 'forgiving wording', the 'everybody' approach, the 'other people' approach, or some combination of these approaches, which have all been first introduced by Barton (1958). With 'forgiving wording', questions are framed in a way that the respective sensitive behavior is forgiven or justified, thereby supposedly encouraging respondents to answer more truthfully. Under the 'everybody' approach, which is intimately related to the 'forgiving wording' approach, the interviewees are told that it is common for many people to show the sensitive behavior ("everybody does it"); the second part of the question then asks if the respondents themselves engage(d) in the respective behavior. Under the 'other people' approach, respondents are asked if they know anybody, possibly with similar characteristics compared to themselves, who engages or engaged in the respective behavior, or to refer to other but similar households / firms in their answers.

In some questionnaires of the World Bank Enterprise Surveys (WBES), a combination of some variant of the 'other people' approach and the use of forgiving wording has been used to obtain information about tax evasion. The specific question asked is: "Recognizing the difficulties many enterprises face in fully complying with taxes and regulations, what percentage of total sales would you estimate the typical firm in your area of activity keeps 'off the books'?". The corresponding variable has been used in research on the determinants of informality; see for instance Dabla-Norris et al. (2008).[7] However, these approaches have in common that respondents may not alter their behavior in terms of truth-telling in response to changes in framing, and

---

[7] Alm and McClellan (2012) take a different approach to estimate tax evasion among firms using data from the World Bank Enterprise Surveys. In the first step, they estimate the determinants of seeing taxation as obstacle; the residual is then used as a measure of unobserved tax morale. In the second step, this measure is used as a determinant of the estimated reporting share of similar firms under the 'other people' approach.

they may realize that the 'other people' approach is used to make inferences about their own behavior.[8] Given the widespread use of the WBES, it is this approach to ask firms questions about tax evasion which serves as our benchmark to assess the merits of a more elaborated method, namely the crosswise model (CM). By contrast, directly questioning respondents about tax evasion has largely been discarded and is likely to result in even larger biases.[9]

2.2 Protection of the Privacy of the Individual Responses

More sophisticated approaches aim at protecting the privacy of the individual answers of the interviewee, while still allowing to compute the prevalence of the illicit or otherwise sensitive behavior in question which is the parameter of interest. Obviously, these approaches do not allow for estimating the determinants of illicit behavior at the individual level in the spirit of Dabla-Norris et al. (2008). However, they address the source of the bias in answers to questions about sensitive and/or illicit activities in the sense that the interviewer does not know whether the respondent engages in illicit or otherwise sensitive activities or not. There are only few papers that apply these methods to tax evasion.

Sealed Envelope Technique

Under the sealed envelope technique first presented by Barton (1958), respondents fill out a questionnaire which is then sealed in an envelope and handed over to the interviewer so that the latter is supposedly unaware of the respondent's answers. However, this technique may relatively easily be manipulated in the sense that it is possible that the individual responses of the interviewees are revealed after the interview. This in turn may prompt interviewees to not answer truthfully.

Randomized Response Technique

The randomized response technique (RRT) is a more prominent and compelling method to protect the privacy of the individual answers of the respondents and dates back to Warner (1965). Under the original RRT set-up, respondents face two statements on a sensitive topic (e.g. "I have cheated on an exam" and "I have not cheated on an exam"). A randomization device such as a dice or the flipping of a coin determines which question the respondent answers, but the

---

[8] This was indeed suggested by interviewer feedback from the pilot of the survey used in this paper.

[9] Obviously, these techniques may be combined with alternative interviewing modes including mail or online surveys. However, business surveys in developing countries are typically carried out using face-to-face interviews to encourage participation, given that questionnaires may be demanding for some respondents. In addition, even for online or mail surveys, manipulation is easily feasible through the interviewers which the respondents may be aware of.

outcome from the randomization device is not revealed to the interviewer. The only information revealed to the interviewer is therefore the respondent's answer ("yes" or "no"), but not which statement or question the answer refers to. This implies that it is not possible to estimate the determinants of tax evasion at the individual level. However, given that the probability distribution of different outcomes of the randomization is known, it is possible to estimate the prevalence of the illicit behavior across the full population. There are various other versions of RRT which we review in the Appendix. However, all RRT survey designs have the general and inherent drawback that respondents have always access to an obvious self-protective answering strategy which they may choose regardless of outcome of the randomization device and the question asked in addition to non-response (Yu et al., 2008; Jann et al., 2012). The crosswise model which we present in detail below avoids this drawback which is the main reason why we argue that this is the optimal approach to study tax evasion in business surveys.

Item Count Technique

Under the item count technique (also referred to as the unmatched count technique, the unmatched block design, or block total response) introduced by Raghavarao and Federer (1979) and described by Droitcour et al. (1991), the sample is split into two groups. Both groups are presented a list of several types of behavior and are asked to state how many of them apply to them. The lists are identical, except that the list presented to one group of respondents also includes the sensitive behavior. The difference in the mean number of items reported by both groups is the estimated proportion reporting the sensitive behavior. While Coutts and Jann (2011) show that the Item Count Technique performs better compared to a number RRT designs, it nevertheless has two disadvantages. On the one hand, even if the list includes innocuous items with a low prevalence, there may still be respondents to whom all types of behavior including the sensitive one apply so that his/her privacy is no longer protected (Chaudhuri and Christofides, 2007). On the other hand, the sample needs to be split so that a larger overall sample is required, but only one subsample provides information about the sensitive behavior (Kuha and Jackson, forthcoming).

Empirical Applications to Assess Tax Evasion

There are only a few papers that apply some of these techniques to analyze tax evasion based on surveys among individuals contrary to our paper which is based on firm-level data. Gërxhani (2007) uses the sealed envelope approach and finds that almost 39% of the respondents in a survey among Albanian households evade personal income tax. Similarly, using data from a survey among German individuals which likewise applies the sealed envelope technique, Becker and Günther (2004) study the determinants of tax evasion, but do not report the prevalence of tax evasion. The analysis of Houston and Tran (2001) may be indeed closest to this paper, even though important differences remain. Using a mail questionnaire to survey Australian households, Houston and Tran (2001) compare both response rates to questions about tax evasion

7

and the share of individuals reporting tax evasion using the DQ and the RRT approaches. An earlier, similar study is Aitken and Bonneville (1980) as part of a taxpayer survey in the U.S. which is summarized by Boruch (1989). They compare the RRT to the sealed envelope method and find that the RRT yields higher estimated levels of tax evasion. Himmelfarb and Lickteig (1982) compare the RRT to a DQ approach in a survey among students. They find that a significantly higher proportion of respondents admit to have falsified their income tax report when the RRT is applied. Musch et al. (2001) use an online survey to study the prevalence of tax evasion among German individuals who self-selected themselves for participation. They compare tax evasion rates under the DQ and an augmented RRT approach which addresses the problem that respondent may adopt a self-protective answering strategy. They find significant differences between the DQ method and their specific RRT approach. However, their survey design requires three subsamples and therefore a larger overall sample so that we do not discuss their approach further.[10]

Table 1: Methods to elicit sensitive information in surveys

Method	Original Paper	Application to Tax Evasion	Main drawback
Forgiving wording and the everybody approach	Barton (1958); see Näher and Krumpal (2012) for a summary	World Bank Enterprise Surveys	No privacy protection
The other people approach	Barton (1958)		
Sealed-envelope technique	Barton (1958)	Becker and Günther (2004), Gërxhani (2007)	Privacy protection mechanism subject to manipulation
Randomized response technique / Warner's method	Warner (1965); see Lensvelt-Mulders et al. (2005) for a meta-analysis	Houston and Tran (2001), Musch et al. (2001), Himmelfarb and Lickteig (1982), Aitken and Bonneville (1980)	Respondents have always access to an obvious self-protective answering strategy which they may choose irrespective of the true answer
Randomized response technique / unrelated question	Horvitz et al. (1967) and Greenberg et al. (1969)		
Randomized response technique / forced alternative	Boruch (1971); see Tourangeau and Yan (2007) for a summary		
Randomized response technique / Benford's law	Diekmann (2012)		

---

[10] There are also papers that apply RRT to other tax-related questions; see for instance Larkins et al. (1997).

Item count technique	Raghavarao and Federer (1979); see Droitcour et al. (1991) for a summary	n/a	Privacy protection not necessarily guaranteed
Crosswise model	Yu et al. (2008), Jann et al. (2012)	this paper	Non-response (as in the case of all other strategies) may still serve as (the only) possible self-protective strategy

2.3   Crosswise Model

In this paper, we use a more recent and arguable better-suited method, namely the crosswise model (CM) introduced by Yu et al. (2008). Similarly to the RRT, respondents are asked two questions, one about a sensitive characteristic denoted by *X* with an unknown distribution, and one about a non-sensitive characteristic denoted by *Y* with a known distribution.[11] If respondents have the sensitive characteristic, $X = 1$ applies, whereas $X = 0$ applies in the opposite case. The same holds for the non-sensitive question, i.e. $Y = 1$ for the non-sensitive characteristic, and $Y = 0$ otherwise. In the survey, respondents face only two options to jointly answer both questions, namely either (1) "*no* to both questions, or *yes* to both questions", or (2) "*yes* to one of the questions, and *no* to the other one". This type of bundling of answers ensures that the CM does not provide respondents with an obvious self-protective strategy which they may resort to if they distrust the questioning strategy. In other words, neither option (1) nor option (2) unambiguously negates both the non-sensitive question and the sensitive question. This is the central advantage of the CM compared to the RRT. Obviously, another self-protective strategy, namely not answering at all is still feasible, but as we show later, this strategy was not chosen in our survey. One reason may be the possibility that respondents may think that non-response could raise the suspicion that they engage in the sensitive behavior; in this sense responding is a 'safer' strategy in the context of the crosswise model.

With two types of characteristics, *X* and *Y*, that each have two values, respondents may be divided in four different subgroups as summarized in Table 2. The prevalence of Y, $p = \Pr(Y = 1)$, is essentially set by the researcher through asking an appropriate non-sensitive question about a characteristic with known distribution. With $p = 1$, the CM (like the RRT) becomes identical to the DQ approach (see Warner, 1965).

---

[11] We follow the notation of Yu et al. (2008).

Categories		Non-sensitive characteristic Y		
		$Y = 0$	$Y = 1$	All
Sensitive characteristic X	$X = 0$	$(1-\pi)(1-p)$	$p(1-\pi)$	$1-\pi$
	$X = 1$	$\pi(1-p)$	πp	π
	All	$1-p$	p	1

Source: Adapted from Yu et al. (2008)

By contrast, the unknown prevalence of the sensitive characteristic X, $\pi = \Pr(X = 1)$, is the parameter of interest and may be estimated using Maximum Likelihood based on the empirical distribution of the answers; below we summarize the proof by Yu et al. (2008). Table 3 (second row) illustrates how the respondent categories summarized in Table 2 map into answer options (1) and (2). It shows that the respondents can always be sure that interviewer does not know the true answer on the sensitive question, irrespective of whether the respondents choose answer option (1) or (2) because the corresponding subgroups are all non-sensitive. The third row contains the unobserved probability of each answer category derived from the corresponding cells of Table 2. The observed number of respondents choosing option (1) is denoted by n_1 which follows a binomial distribution with $E(n_1) = n\lambda$ and $Var(n_1) = n\lambda(1-\lambda)$. The log Likelihood function is

$$\log L = n_1 \log[(1-\pi)(1-p) + \pi p] + (n-n_1)\log[\pi(1-p)p(1-\pi)] \tag{1}$$

which is analogous to the likelihood function in the CM presented by Yu et al. (2008) and similar to those under the RRT presented by Warner (1965). The first-order condition is

$$\frac{n_1(2p-1)}{\pi p + (1-\pi)(1-p)} = \frac{(n-n_1)(2p-1)}{\pi(1-p) + p(1-\pi)} \Leftrightarrow (1-\pi)(1-p) + \pi p = \frac{n_1}{n} \tag{2}$$

Solving (2) for π allows us to obtain the Maximum Likelihood estimator of π, $\hat{\pi}$:

$$\hat{\pi} = \frac{(\hat{\lambda} + p - 1)}{(2p - 1)} \tag{3}$$

where $\hat{\lambda} = \dfrac{n_1}{n}$ is the Maximum Likelihood estimate of λ from (2) and thus equal to the observed share of respondents choosing answer option (1).

Table 3: Answer options and probabilities

Answer option	(1) *"no to both questions, or yes to both questions"*	(2) *"yes to one of the questions, and no to the other one"*
Respondent categories	$\{X = 0 \cap Y = 0\} \cup \{X = 1 \cap Y = 1\}$	$\{X = 1 \cap Y = 0\} \cup \{X = 0 \cap Y = 1\}$
Unobserved probability	$\lambda = (1-\pi)(1-p) + \pi p$	$1 - \lambda = \pi(1-p) + p(1-\pi)$
Observed number of respondents	n_1	$n - n_1$
Distribution function of n_1	Binomial (n, λ) $E(n_1) = n\lambda$, $Var(n_1) = n\lambda(1-\lambda)$	
Estimated probability	$\hat{\lambda} = \dfrac{n_1}{n}$	$1 - \hat{\lambda}$

Source: Own compilation based on Yu et al. (2008)

The variance of $\hat{\pi}$ can be written as

$$Var(\hat{\pi}) = Var\left(\frac{\hat{\lambda} + p - 1}{2p - 1}\right) = Var\left(\frac{\hat{\lambda}}{2p - 1}\right) = \frac{Var\left(\frac{n_1}{n}\right)}{(2p-1)^2} = \frac{Var(n_1)}{n^2(2p-1)^2} = \frac{n\hat{\lambda}(1-\hat{\lambda})}{n^2(2p-1)^2} = \frac{\hat{\lambda}(1-\hat{\lambda})}{n(2p-1)^2} \quad (3)$$

In order to obtain an unbiased estimate of (3), we have to multiply (3) by $n/(n-1)$, given that the mean of the distribution function is unknown (Bessel's correction):

$$\overline{Var}(\hat{\pi}) = \frac{\hat{\lambda}(1-\hat{\lambda})}{(n-1)(2p-1)^2} \quad (4)$$

As equation (4) demonstrates, increasing the level of privacy protection lowers the efficiency of the estimate, as decreasing p raises the variance as well. With $n \to \infty$, (4) allows constructing confidence intervals for $\hat{\pi}$:

$$\hat{\pi} = \pm Z_{\alpha/2}\sqrt{\overline{Var}(\hat{\pi})} \quad (5)$$

where $Z_{\alpha/2}$ denotes the $((1-\alpha/2)\times 100)th$ percentile of the standard normal variable Z.

The study by Jann et al. (2012), which compares the estimated prevalence of plagiarism under CM and DQ, is the only application of this approach so far. The students participating in the survey were asked about past instances of partial and severe plagiarism in assignments such as

seminar and term papers. With respect to the unrelated question, Jann et al. (2012) asked whether the month of birth of a close family member (i.e., father in one case, and mother in another case) is January, February or March. Assuming that months of births are equally distributed, the prevalence of *Y* in this case would be 0.25. For partial plagiarism, the CM yielded a significantly higher prevalence estimate than DQ, with a difference amounting to 15 percentage points (7.3% vs. 22.3%). In case of severe plagiarism, the estimates remained low for both approaches (1.0% for DQ and 1.6% for the CM, respectively), and the difference was insignificant. The results therefore suggest that enhanced anonymity of CM is at least in some cases important for the results.

3   Data and Survey Design

In this paper, we hypothesize that the prevalence of tax evasion estimated using information from business surveys is higher under the CM approach compared to our benchmark approach, and that these differences arise due to different levels of privacy protection. Our benchmark approach is the standard one used to study illicit behaviour of firms in business surveys such as the World Bank Enterprise Surveys for instance, namely a combination of the 'other people' approach and the use of forgiving wording. Compared to Jann et al. (2012), our benchmark may be expected to yield higher estimates of the share of respondents showing the sensitive behavior compared to direct questioning.

In order to test this hypothesis, we make use of novel information from a rich survey among small and medium firms in Serbia carried out in November and December 2012 on behalf of the GIZ Public Finance Reform Project[12]. The survey focuses on the perceived efficiency and customer orientation of the Serbian tax system and administration, and on different aspects of firms' tax compliance behaviour, as well as on firms' attitudes towards paying taxes. It was implemented by a professional survey company, using face-to-face interviews with firms. The privacy of responses by the interviewees was guaranteed by GIZ as an honest broker. In addition, the survey company is not associated as having any connection with the Serbian government or the Serbian revenue administration.

The survey covers 422 firm-level observations, and the sample was drawn from the Serbian Business Register (2011). It is representative of micro, small and medium-sized Serbian firms with 1 to up to 99 employees operating in manufacturing and service sectors.[13] In line with standard practice in business surveys, we excluded agriculture and fishing. Given the focus on taxation, we also excluded firms operating in the mining and quarrying sectors, in financial in-

---

[12] Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH is a German aid agency.

[13] The survey is based on the NACE Rev. 1.1 industry classification.

termediation as well as real estate and renting because the nature of these firms and/or the tax regime they are subject to differs from other firms, complicating comparisons. We also excluded firms operating in various business service sectors (NACE 73 and NACE 74), again, because the nature of these firms may differ significantly from other firms in our sample.[14] Finally, given that we are interested in private sector firms, we exclude Section L (public administration and defence, compulsory social security), Section M (education), Section N (health and social work) and Section O (Other community, social and personal service activities), as these activities are likely to be carried out to a significant extent by public entities or state-owned enterprises. Table 4 provides a summary of the sample.

The sample was split into two subsamples which do not show any significant differences in terms of size-sector-region structure with an almost identical number of firms in each subsample (see Table 4).[15] In the beginning of the interview, firms were told that the objective of the survey is to better understand the problems and obstacles that small and medium firms face, in particular in the area of taxation to make tax policy and tax administration more business friendly.

Similar subsamples ensure that estimated differences in tax evasion between both groups are most likely not due to differences in actual tax compliance behaviour, but rather due to differences in truth-telling. In order to achieve this, within each size-sector-region strata, firms were randomly allocated to each of the subsamples. Tax evasion within the first subsample, referred to as the 'benchmark group', was then estimated using the 'other people' approach in combination with forgiving wording, whereas tax evasion in the second subsample, referred to as the 'treatment group', was estimated using the crosswise model approach.

---

[14] This exclusion restriction covers for instance accounting and tax advisory firms, among others. The views on taxation of managers of these types of firms are likely to differ fundamentally from other firms, and their managers may answer more strategically in questions about tax evasion, reflecting the 'self-interest' of these sectors.

[15] One reason for the small difference in terms of subsample size relates to differences in participation rates in the survey. However, participation in the interview did not depend on whether firms were assigned to the benchmark group or the treatment group as this information was only revealed to them in the course of the interview.

	Number of obs.	Distribution (in %)			
		Benchmark		Crosswise	
Firm size		*weighted*	*unweighted*	*weighted*	*unweighted*
1-4 employees	222	66.49	50.93	66.50	54.37
5-19 employees	112	25.78	27.78	25.78	25.24
20-99 employees	88	7.73	21.30	7.72	20.39
Sector					
Manufacturing	117	25.00	27.31	25.00	28.16
Retail and wholesale	198	49.96	44.91	49.96	49.03
Other services	107	25.04	27.78	25.04	22.82
Region					
Belgrade	162	41.15	39.35	41.15	37.38
Vojvodina	142	33.15	33.80	32.97	33.50
Central Serbia	118	25.70	26.85	25.87	29.13
Total (number of obs.)	422	216		206	

Source: own compilation

We distinguish two modes of tax evasion, sales underreporting and envelope wages, which are both very common in Serbia. Since the CM only allows asking dichotomous questions, we estimate the share of firms that underreport at least 10% of actual sales and wages, respectively. We chose the 10%-threshold because we consider underreporting above this threshold as significant from an economic point of view. Under the benchmark approach, the following question was asked to estimate the extent of sales underreporting (an analogous question was asked to estimate the extent of envelope wages):

"Firms often struggle to meet all tax obligations which impose a significant burden on firms. According to your experience and judgment, do firms like this underreport at least annual 10% of annual sales to STA for VAT and/or profit tax?"

The respondents of the treatment group received an introduction to the crosswise model explaining that this questioning technique is designed to protect the privacy of their answers while acknowledging that this may seem strange to them. Interviewees are unlikely to understand the exact mechanism of interfering tax underreporting from the CM design; nevertheless, given the questions and the answer options, it does seem likely that they understand that the privacy of their particular answer is protected. In the specific context of this survey, the non-sensitive characteristic used by Jann et al. (2012), namely the birth month of one parent, may be subject to the criticism that respondents do not know the birth months of the parents (e.g. if the respondent is an orphan or has been abandoned by one parent as a result of past armed conflicts). Alternatively, respondents may be afraid that their parents' month of birth can in principle be obtained by the interviewer through official records so that the privacy of the responses

is no longer protected. In addition, we did not have access to statistics that could be used to estimate the distribution of birth months.

Therefore, contrary to Jann et al. (2012), we chose the last digit of the best friend's mobile phone number as the non-sensitive characteristic and asked if it is "0 or 1" and "8 or 9", respectively. Our assumption is that the distribution of the last digit of mobile phone numbers is uniform giving rise to *p=0.2*. Our results are robust to changes in the underlying distribution and in particular to higher prevalence rates which we discuss below, given that the allocation of mobile phone numbers by the operators is not known with certainty.

Nevertheless, we believe that the allocation of mobile phone numbers, in contrast to landline numbers, can be expected to be done in a rational and well-defined manner even in developing countries which leads to a uniform distribution of the last digits so that Benford's law which we discuss in the Appendix does not apply. The reason is that mobile phone operators that are well-run firms often owned by multinational enterprises are in charge of the number allocation in Serbia.[16] Two allocation principles are feasible. Under the first one, mobile numbers would be allocated in a piecemeal fashion, where a new phone number is simply the highest existing number plus 1. Under an alternative approach, mobile phone numbers are allocated randomly. Given the large number of mobile phone subscriptions, both approaches would imply a nearly equal distribution of the last digit. Obviously, there may be preferred numbers allocated to special customers where the last couple of digits are all identical. However, such occurrences can be expected to be rare thereby not significantly affecting the overall distribution of the last digits of the phone numbers.[17]

Specifically, under the crosswise model, we asked the following two questions simultaneously to estimate the extent of sales underreporting:

"Is the last digit of your best friend's phone number / of the number of the person you call most often 0 or 1? Does your firm underreport at least 10% of your annual sales to STA for VAT and/or profit tax?"

In order to estimate the extent of envelope wages, the following two questions were posed simultaneously:

---

[16] The Republic Agency for Electronic Communications in Serbia allocates the entire range of subscriber numbers which follow the network code to the mobile phone operator. This implies that the mobile phone operator is also in charge of allocating the last digit of mobile phone numbers to the individual subscribers. See Serbian Numbering Plan for more details (http://www.ratel.rs/upload/documents/Regulativa/Plan_numeracije/Numbering_plan.pdf).

[17] Obviously, there may be other factors that potentially confound this randomization mechanism. Some participants may only know the landline number or imperfectly remember the mobile number, even though participants were asked to look up the number in their phone. However, these factors are unlikely to be systematic and/or of significance.

"Is the last digit of your best friend's phone number / of the number of the person you call most often 8 or 9? Does your firm pay more than 10% of the total wage bill in cash to avoid wage tax and social security contributions?"

We then compare estimates of both sales underreporting and envelope wages obtained through forgiving wording as our benchmark which is commonly used to elicit sensitive information in business surveys and through applying the crosswise model. We hypothesize that estimated tax evasion is higher under the CM approach because it offers enhanced privacy protection.

Obviously, differences in estimated tax evasion across both groups may also arise due to minor difference in the framing and/or the design of the questions, but here we argue that this is highly unlikely. On the one hand, respondents within the benchmark group are asked to refer to firms similar to their own which may however differ in terms of tax compliance behavior, or respondents do not use tax non-compliance of their own firm to estimate tax non-compliance of similar firms. However, in the pilot study preceding the actual survey, some respondents even told the interviewers that they understand that this question is used to infer their own tax non-compliance behavior, and their own behavior is a natural reference point to estimate the behavior of similar firms.

On the other hand, in the benchmark group, spontaneous non-response was permitted, either because the respondent refused to answer, or because the respondent states that she/he does not know; however, non-response was not read out by the interviewers as a possible answer option. We assume that non-responses are equivalent to 'no' answers to the tax evasion questions, given that non-response is a common strategy to avoid admitting illicit or otherwise sensitive behavior in surveys. Overall, about 23% of the respondents followed the self-protective strategy with respect to the underreporting of sales, and about 22% for the respective question on envelope wages. By contrast, in the treatment group, spontaneous non-response was not permitted in the sense that interviewers 'pushed harder' to obtain answers (but ultimately, non-response was still feasible). Here, non-response is more likely to result from 'laziness to participate' in the crosswise model which may have seemed odd to respondents, rather than from the reluctance to answer the questions. Yet, no respondent chose to refrain from answering the sensitive questions asked under the CM. In a robustness check, we test whether this minor difference in the survey design between both groups on its own gives rise to differences in our estimates of tax evasion between both groups.

4 Results

4.1 Baseline Results

Specification (1) in Table 5 contains our baseline results. Here, tax evasion is estimated using sampling weights to ensure that the results are representative. The results show that the estimated share of firms that underreport at least 10% of sales amounts to 23.9% of the firms under the FW, and to 33.9% of the firms under the CM. With respect to the payment of wages in cash, the analogous estimated share is 26.5% under FW, which is comparable to the underreporting of sales. Again, the estimate is larger (30.2%) if we apply the CM. The CM estimates are obtained using equation (3)

From equation 5, we construct confidence intervals to check the significance of the prevalence estimates; none of the 99% confidence intervals includes zero, which means that the prevalence estimates for both methods are highly significant on the 99% level. To further evaluate the statistical significance of the differences between the estimates, we apply a two-sample proportion test. The z-value is estimated by dividing the difference between the FW- and the CM estimate by the joint standard error. For the underreporting of sales, the difference between FW and CM is statistically significant at the 10% level (the p-value is 0.059). By contrast, we find no significant differences for estimates of the prevalence of envelope wages (the p-value is 0.167). Our results therefore partially confirm our hypothesis, namely that tax evasion estimates are higher under the CM approach given that it is designed in a way that the privacy of individual responses are protected.

Nevertheless, the difference in terms of significance between sales and wage underreporting requires interpretation. The statistical interpretation is that the difference between estimates obtained under FW and CM may be insignificant because under the CM approach, the increase in anonymity through the non-sensitive question also introduces an additional source of error. To compensate for this, a larger sample size or a lower level of anonymity would be necessary. However, large samples in business surveys are usually expensive, especially for face-to-face interviews. Lowering the level of anonymity to compensate for the higher variance due to the sample size decreases chances that respondents answer truthfully or participate at all. Obviously, one can only speculate whether a larger sample size or lower levels of anonymity would have rendered the estimated difference of the prevalence of wage underreporting between FW and CM significant.

Alternatively, there may be economic factors as well which explain why the CM does not yield higher and statistically different results compared to the benchmark approach. On the one hand, respondents may believe, at least to some extent, that wage underreporting attracts less severe penalties in case it is revealed to authorities, irrespective of whether this is the case. In

the survey, respondents were also asked to rate the severity of penalties associated with different taxes and social contributions. Indeed, 28.67% of the respondents consider penalties for VAT evasion as most severe, which is larger than the corresponding shares of respondents for all other taxes and social contributions. On the other hand, respondents from the benchmark group may have believed that information on wage underreporting is less useful for the tax administration if it is revealed. In the latter case, in order to impose penalties, the tax administration would probably have to start a formal investigation, rather than imposing penalties simply based on the information provided through the survey. Contrary to sales underreporting, which can be detected relatively easily by the authorities, for instance through checking if customers of a particular firm were given correct invoices, investigating whether a particular firm pays envelope wages is more demanding. The reason is that both, employers and employees often have strong incentives to hide this practice. Contrary to this, customers are likely to be indifferent. Both factors imply that protecting the privacy of the individual responses is more important for questions addressing the extent of sales underreporting compared to questions addressing the extent of envelope wages. This may explain why we do not find significant differences between benchmark and CM estimates with respect to the prevalence of envelope wages.

4.2   Robustness Checks

We test the robustness of our results in several ways – see the remaining specifications in Table 5. First, given that most other papers that apply RRT and CM approaches do not use sampling weights, we re-assess the difference between benchmark and CM estimates using no weights (specification 2). The difference for estimated sales underreporting remains positive and significant, although it slightly decreases.

Second, we only include micro firms with up to 4 employees in our sample (specification 3) which almost halves the sample (in total, there are 222 firms left). While we recognize that further limiting the sample size increases the variance of the CM estimates, it is still insightful to exclude larger firms. The latter are likely to have access to sophisticated, legal tax avoidance strategies, and they are often monitored more intensively by the authorities, especially in developing countries, so that estimates of tax evasion that exclude large firms are likely to be larger. This also implies that the protection of the privacy of the answers of the managers of these firms is less important. The difference between the benchmark and CM estimates of sales underreporting indeed increases to 0.17 and is significant at the 5% level even though the sample size decreases. By contrast, the results hardly change with respect to wage underreporting.

Third, we check whether allowing for spontaneous non-response affects the results (specification 4). Our benchmark question differs in the sense that interviewers accepted spontaneous non-response and push respondents 'less hard' to provide an answer compared to the treat-

ment group, where respondents were 'pushed harder' to answer. In this specification, we use an alternative question to estimate the extent of tax evasion firms under the benchmark approach which was posed immediately after the questions on sales and wage underreporting used in the baseline specification. With respect to sales underreporting, the question is: "According to your experience and judgment, on average, what percent of total annual sales do firms like this one underreport to STA for VAT and/or profit tax?" With respect to wage underreporting, the question is: "On average, what share of wages do firms like this one typically pay in cash?" The questions do not use forgiving wording, but the questions preceding each of these questions do implicitly justify tax evasion.

For both questions, respondents could select one of the following options, 1) 0%, 2) 1% - 10%, 3) 11% - 25% or 4) More than 25%. While for simplification, this scale was chosen to be not continuous in the survey, we still assume that the combined share of respondents selecting options 3) and 4) corresponds to the share of respondents that underreport at least 10% of wages and sales, respectively. Contrary to the questions asked to the benchmark group firms in the remaining specifications, this question did not allow for spontaneous non-response, similarly to the questions asked under the CM approach, where interviewers did not accept spontaneous non-response, but pushed harder to obtain a response. Interestingly, the difference between the benchmark and the CM estimates of sales underreporting are again much higher compared to the baseline specification and highly significant at the 1% level, and the difference between the benchmark and CM estimates of wage underreporting also increase and are likewise significant at the 5% level.

Finally, one important concern may be related to our assumption about the distribution of the last digits of mobile phone numbers in Serbia. In particular, the results of the CM also depend on the ex-ante chosen probability that the last digit of the best friend's mobile phone number is 0 or 1 (and 8 or 9, respectively). So far, we have assumed a uniform distribution of the last digits. However, it could be argued that in fact, the probability that the last digit is '0' or '1' is higher than 20%, and that the probability that the last digit is '8' or '9' is below 20%. To address this, in specification 5, we reevaluate the difference between the benchmark and CM estimates of sales and wage underreporting where we assume that the probability is 10% higher (lower) that the last digits are '0' or '1' ('8' or '9'). The difference between the benchmark and the CM estimate (0.09) is only marginally smaller than under specification 1 when it comes to sales underreporting, and it remains significant at the 10% level even if we assume that the probability increases to $p=0.22$. We likewise find no evidence that a lower probability that the last digits are '8' or '9' ($p=0.18$) affects the estimates for wage cash payments.

Table 5: Results

Specification	Mode of tax evasion	Benchmark estimates	Crosswise model (CM)	Difference
(1)	Underreporting of sales (at least 10% of total sales)	0.24 (0.03)	0.34 (0.06)	0.10* (0.06)
	Paying wages in cash (at least 10% of wages)	0.27 (0.03)	0.30 (0.06)	0.04 (0.06)
(2)	Underreporting of sales (at least 10% of total sales)	0.24 (0.03)	0.33 (0.06)	0.09* (0.06)
	Paying wages in cash (at least 10% of wages)	0.26 (0.03)	0.33 (0.06)	0.07 (0.06)
(3)	Underreporting of sales (at least 10% of total sales)	0.21 (0.04)	0.38 (0.08)	0.17** (0.09)
	Paying wages in cash (at least 10% of wages)	0.25 (0.04)	0.31 (0.08)	0.06 (0.09)
(4)	Underreporting of sales (at least 10% of total sales)	0.17 (0.03)	0.34 (0.06)	0.17*** (0.06)
	Paying wages in cash (at least 10% of wages)	0.17 (0.03)	0.30 (0.06)	0.13** (0.06)
(5)	Underreporting of sales (at least 10% of total sales)	0.24 (0.03)	0.33 (0.06)	0.09* (0.07)
	Paying wages in cash (at least 10% of wages)	0.26 (0.03)	0.31 (0.05)	0.05 (0.06)

Source: own compilation

Notes: Standard errors in parenthesis; significance levels: * $p<0.1$, **$p<0.05$, ***$p<0.01$

(1) full sample; weighted proportions

(2) full sample; unweighted proportions

(3) only micro firms with 1-4 employees included; weighted proportions

(4) full sample; weighted proportions; under FW, no spontaneous non-response accepted

(5) full sample; weighted proportions; alternative assumptions about prevalence of non-sensitive characteristic

5 Conclusions

This paper uses new evidence from a business survey in Serbia to measure the extent of tax evasion. Obviously, respondents in such surveys can be expected to have strong incentives to not answer truthfully to questions about sensitive topics. We therefore employ a new survey method, referred to as the crosswise model, that protects the privacy of respondents through bundling of questions about tax evasion and 'harmless' topics and that has been successfully applied elsewhere in the social sciences. Contrary to other surveys that examine tax evasion, we differentiate between two types of tax evasion, namely underreporting of sales when the crosswise model is used and envelope wages. We show that a significantly higher share of managers of small and medium size firms admits considerable underreporting of sales compared to conventional approaches. The result is robust to a number of alternative specifications. We con-

clude that such an approach delivers a more, though possibly not fully, realistic picture of tax evasion.

One type of criticism that may apply to our survey design relates to the size of the sample. Ultimately, based on our findings, it is difficult to judge whether the crosswise model (CM) does indeed not yield statistically different estimates from our benchmark, or whether this result is driven by the limited sample size and the large variance that comes with it. The size of our sample was dictated by budget constraints, and given that our data is part of a detailed face-to-face business survey about tax issues, the size of our overall sample is appropriate. In addition, to study the relative merits of the crosswise model compared to our benchmark, we had to split the sample which further reduced the number of observations. Ideally, future research would therefore carry out a similar survey with a larger sample.

Our results are unlikely to be affected by any bias relating to reluctance to answer truthfully, and this study is the first attempt to obtain credible and more detailed picture about the extent of tax evasion from businesses who, from a revenue perspective, are the most important tax-payers as they remit the bulk of taxes to revenue authorities. Nevertheless, the crosswise model raises its own issues. One central one is that by definition, this approach does not generate data that allow studying the determinants of tax evasion at the individual level. Another one that could be addressed in future research is that currently, the crosswise model only allows obtaining dichotomous information about tax evasion. Unlike topics such as cheating in exams or drug abuse where a dichotomous answer is revealing and informative, more detailed quantitative information about the extent of tax evasion is crucial. In other words, from a policy perspective, it is also important to obtain the precise estimate of tax evasion and hence of foregone revenue. The most obvious methodological extension which we leave for future research would therefore be to amend the CM to obtain this type of quantitative information as well, for instance through asking several dichotomous questions.

6 References

Aitken, S. S., & Bonneville, L. (1980). A General Taxpayer Opinion Survey. Prepared for the Office of Planning and Research, Internal Revenue Service. Washington: CSR, Inc.

Alm, J. (2012). Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies. International Tax and Public Finance, 19(1), 54–77.

Alm, J., & McClellan, C. (2012). Tax Morale and Tax Compliance from the Firm's Perspective. Kyklos, 65(1), 1–17.

Almunia, M., & Lopez-Rodriguez, D. (2012). The efficiency cost of tax enforcement: evidence from a panel of Spanish firms. unpublished manuscript.

Barton, A. H. (1958). Asking the Embarrassing Question. Public Opinion Quarterly, 22(1), 67.

Becker, R., & Günther, R. (2004). Selektives Antwortverhalten bei Fragen zum delinquenten Handeln : eine empirische Studie über die Wirksamkeit der "sealed envelope technique" bei selbst berichteter Delinquenz mit Daten des ALLBUS 2000. ZUMA Nachrichten, 28(54), 39–59.

Boruch, R. F. (1971). Assuring confidentiality of responses in social research: a note on strategies. The American Sociologist, 6(4), 308–311.

Boruch, R. F. (1989). Appendix B: Experimental and Quasi-Experimental Designs in Taxpayer Compliance Research. In J. A. Roth, J. T. Scholz, & A. D. Witte (Eds.), Taxpayer compliance - Volume 1: An agenda for research (Vol. 03, pp. 338 – 379). University of Pennsylvania Press.

Chaudhuri, A., & Christofides, T. C. (2007). Item Count Technique in estimating the proportion of people with a sensitive feature. Journal of Statistical Planning and Inference, 137(2), 589–593.

Coutts, E., & Jann, B. (2011). Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT). Sociological Methods & Research, 40(1), 169–193.

Dabla-Norris, E., Gradstein, M., & Inchauste, G. (2008). What causes firms to hide output? The determinants of informality. Journal of Development Economics, 85(1-2), 1–27.

Diekmann, A. (2012). Making Use of "Benford's Law" for the Randomized Response Technique. Sociological Methods & Research, 41(2), 325–334.

Droitcour, J., Caspar, R.A., Hubbard, M.L., Parsley, T.L., Visscher, W. & Ezzati, T.M. (1991). The item count technique as a method of indirect questioning: A review of its development and a case study application. Measurement errors in surveys, 185–210 Wiley Online Library.

Edgell, S. E., Himmelfarb, S., & Duchan, K. L. (1982). Validity of Forced Responses in a Randomized Response Model. Sociological Methods & Research, 11(1), 89–100.

Feldman, N. E., & Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. The Economic Journal, 117(518), 327–352.

Gemmell, N., & Hasseldine, J. (2012). The Tax Gap: A Methodological Review. Advances in Taxation, 20, 203–231.

Gërxhani, K. (2007). "Did You Pay Your Taxes?" How (Not) to Conduct Tax Evasion Surveys in Transition Countries. Social Indicators Research, 80(3), 555–581.

Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R., & Horvitz, D. G. (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. Journal of the American Statistical Association, 64(326), 520–539.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. American Economic Review, 102(2), 994–1028.

Hessing, D. J., Elffers, H., & Weigel, R. H. (1988). Exploring the limits of self-reports and reasoned action: An investigation of the psychology of tax evasion behavior. Journal of personality and social psychology, 54(3), 405–413.

Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. Journal of Personality and Social Psychology, 43(4), 710–717.

Houston, J., & Tran, A. (2001). A survey of tax evasion using the randomized response technique. Advances in Taxation, 13, 69–94.

Jann, B., Jerke, J., & Krumpal, I. (2012). Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism. Public Opinion Quarterly, 76(1), 32–49.

Krstić, G., Schneider, F., Arandarenko, M., Arsić, M., Radulović, B., Ranđelović, S., & Janković, I. (2013). *The Shadow Economy in Serbia. New Findings and Recommendations for Reform*. USAID Serbia and Foundation for the Advancement of Economics.

Kuha, J., & Jackson, J. (2013). The Item Count Method for Sensitive Survey Questions: Modelling Criminal Behaviour. Journal of the Royal Statistical Society: Series C (Applied Statistics).

Larkins, E. R., Hume, E. C., & Garcha, B. S. (1997). The validity of the randomized response method in tax ethics research. Journal of Applied Business Research (JABR), 13(3), 25–32.

Lensvelt-Mulders, G. J. L. M. (2005). Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation. Sociological Methods & Research, 33(3), 319–348.

Long, S. B., & Swingen, J. A. (1991). Taxpayer compliance: Setting new agendas for research. Law & Society Review, 25(3), 637–683.

Musch, J., Bröder, A., & Klauer, K. C. (2001). Improving survey research on the World-Wide Web using the randomized response technique. Dimensions of Internet science, 179–192.

Näher, A.-F., & Krumpal, I. (2012). Asking sensitive questions: the impact of forgiving wording and question context on social desirability bias. Quality & Quantity, 46(5), 1601–1616.

Pissarides, C. A., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. Journal of Public Economics, 39(1), 17–32.

Pomeranz, D. D. (2013). No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. Unpublished manuscript. Retrieved from http://dash.harvard.edu/handle/1/10208237

Raghavarao, D., & Federer, W. T. (1979). Block total response as an alternative to the randomized response method in surveys. Journal of the Royal Statistical Society. Series B (Methodological), 40–45.

Schmidt, V., & Vaughan-Whitehead, D. (2011). Wage trends and wage policies in South-East Europe: What impact of the crisis? In V. Schmidt & D. Vaughan-Whitehead (Eds.), The Impact of the Crisis on Wages in South-East Europe (pp. 1 – 26). International Labour Organization.

Schneider, F., & Enste, D. H. (2000). Shadow Economies: Size, Causes, and Consequences. *Journal of Economic Literature*, 38, 77–114.

Slemrod, J., Blumenthal, M., & Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. Journal of Public Economics, 79(3), 455–483.

Slemrod, J. (2007). Cheating Ourselves: The Economics of Tax Evasion. Journal of Economic Perspectives, 21(1), 25–48.

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. International Tax and Public Finance, 19(1), 25–53.

Torgler, B. (2011). Tax morale and compliance : review of evidence and case studies for Europe. World Bank Policy Research Working Paper Series, (5922).

Tourangeau, R., & Yan, T. (2007). Sensitive Questions in Surveys. Psychological Bulletin, 133(5), 859–883.

Warner, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63–69.

Yu, J.-W., Tian, G.-L., & Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristic: design and analysis. Metrika, 67(3), 251–263.

Zitzewitz, E. (2012). Forensic Economics. Journal of Economic Literature, 50(3), 731–69.

7   Appendix: Other Versions of the Randomized Response Technique

There exist various other versions of the RRT designs. First, under the *unrelated question method* (Greenberg et al., 1969), respondents are asked to answer a sensitive and a non-sensitive question that is usually unrelated to the sensitive question. For the unrelated question, the distribution of "yes"-answers can be either known (e.g. birth month) or unknown. In the latter case, the proportion of respondents sharing the non-sensitive characteristic has to be estimated. However, in both instances, the unrelated question method is more efficient than Warner's (1965) model because a "yes" statement under the original RRT either implies that the respondent belongs to the group with the sensitive characteristic, or that the respondent does *not* belong to this group (Greenberg et al., 1969). For a true proportion of 5% sharing the sensitive attribute, Greenberg et al. (1969) show that the variance of the original RRT is more than five times larger than the variance under the unrelated question method with a known distribution of "yes"-answers, and it is still larger even when the distribution of "yes"-answers on the innocuous question is unknown.

The *forced alternative method* represents yet another low-variance version of the original RRT. Depending on the outcome of the randomization device, respondents either have to give a "yes"-answer irrespective of the true answer on the sensitive question, a "no"-answer (again irrespective of the true answer), or they have to answer the sensitive question truthfully (Edgell et al., 1982; Tourangeau and Yan, 2007). Based on post-experimental feedback, Edgell et al. (1982) argue that the method provides a relatively high level of anonymity, but they also stress that participants may feel uncomfortable by being directed to a sensitive question. In a meta-survey on the RRT, Lensvelt-Mulders et al. (2005) find that the forced alternative method is by far the most commonly used version of the RRT.

Finally, Diekmann (2012) proposes another version of RRT that exploits Benford's law to reduce the trade-off between efficiency and privacy protection inherent to RRT. The choice of the probability that the respondent answers the sensitive question is important because a smaller probability decreases the chances to arrive at the sensitive question whereas a larger probability increases privacy protection but also the variance of the estimator. Benford's law states that the empirical distribution of the numbers from 1 to 9 is not uniform in many instances (e.g. the first digits of house-numbers). Instead, the numbers follow a Newcomb-Benford distribution with a probability of roughly 0.7 for the digits 1,2,3,4, and of 0.3 for the remaining ones. The advantage of the method is the 'Benford illusion', i.e. many respondents may believe that the numbers are uniformly distributed with a perceived probability for the first four digits of 0.4 and of 0.6 for the remaining ones. This allows designing surveys in such a way that the perceived privacy protection is higher than the actual one as a means to increase the efficiency of the estimator.

8 Appendix: English Translation of Questionnaire

8.1 Introductory Statement

Good morning/day/evening, my name is _____. I am an interviewer in the research agency [*name of survey company*] which carries out various types of research on different topics, and I would appreciate if you agreed to answer some questions for me. The survey is carried out on behalf of Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), a publically owned German organization for international development cooperation. GIZ supports the Serbian government in implementing reforms that strengthen the Serbian economy. For this reason, GIZ uses this survey to better understand the problems and obstacles that small and medium firms face, in particular in the area of taxation to make tax policy and tax administration more business friendly.

Your firm has been selected for this survey, and your views are very important to GIZ. Participating will offer you the unique opportunity that your views and problems will feed into the policy advice that GIZ gives to the Serbian government. This advice improves the conditions of firms like yours.

Confidentiality is very important for [*name of survey company*] and GIZ. As a German publically owned company, GIZ generally obeys strict international data protection standards. All obtained answers will be treated as group data and will be used solely for the purpose of this survey and not passed on to third parties.

8.2   Benchmark Questions

C22	*[INT] Ask only if split A.*
	Firms often struggle to meet all tax obligations which impose a significant burden on firms. According to your experience and judgment, do firms like this underreport at least 10% of annual sales to STA for VAT and/or profit tax?
		1. Yes
		2. No
		   DK (Do not know – do not read)
		   REF (Refusal – do not read)	

			C23
C23	*[INT] Ask only if split A.* According to your experience and judgment, on average, what percent of total annual sales do firms like this one underreport to STA for VAT and/or profit tax?	1. 0% 2. 1% - 10% 3. 11% - 25% 4. More than 25%	
			C24

C24	*[INT] Ask only if split A.*
	Paying wages in cash is often easier to handle and may be seen as a more immediate reward for their work by employees, and some employees may prefer to get some cash, for instance to avoid withdrawing from banks. As another advantage, this avoids wage taxes and social security contributions. According to your experience and judgment, on average, do firms like this one pay more than 10% of the total wage bill in cash to avoid wage tax and social security contributions?
		1. Yes
		2. No
		   DK (Do not know – do not read)
		   REF (Refusal - do not read)	

			C25
C25	*[INT] Ask only if split A.* On average, what share of wages do firms like this one typically pay in cash?	1. 0% 2. 1% - 10% 3. 11% - 25% 4. More than 25%	
			C26

8.3   Crosswise model questions

[INT] Questions C26-C27 ask only if companies belong to split B!
[INT] Only if split B. [INT] Must read! The following questions may be sensitive to you. To protect your privacy, we apply a new kind of questioning-technique. There are two questions to answer. Please consider carefully how you would answer each of the questions ("Yes" or "No"), but do not say it. Then either choose option A 'Yes' to one of the two questions, and 'No' to the other one or option B 'Yes' to both questions, or 'No' to both questions'. We understand that you may find this technique odd, but it ensures that your privacy will be fully protected as we do not know your specific answers to each of the questions, and we do not know the phone number of your friend. With the help of statistical methods, we can only compute the total number of firms for whom the second question applies. Answering the question truthfully is therefore riskless for you, but helps us tremendously.

C26	*[INT] Only if split B* Think of your best friend's mobile phone number or of the mobile phone number of the person you call most often - please look that number up in your mobile if you do not remember it now and identify the last digit of the phone number. *[INT] Wait for the respondent to look at his mobile or to signal that he knows the last digit* *[INT] If respondent finds this weird, reassure that*	1. 'Yes' to one of the two questions, and 'No' to the other one 2. 'Yes' to both questions, or 'No' to both questions	
			C27

	we ask this question simply to protect his privacy Is the last digit of your best friend's phone number / of the number of the person you call most often 0 or 1? Does your firm underreport at least 10% of your annual sales to STA for VAT and/or profit tax?		
C27	*[INT] Only if split B.* Think of your best friend's mobile phone number or of the mobile phone number of the person you call most often - please look that number up in your mobile if you do not remember it now and identify the last digit of the phone number. *[INT] Wait for the respondent to look at his mobile or to signal that he knows the last digit* *[INT] If respondent finds this weird, reassure that we ask this question simply to protect his privacy* Is the last digit of your best friend's phone number / of the number of the person you call most often 8 or 9? Does your firm pay more than 10% of the total wage bill in cash to avoid wage tax and social security contributions?	1. 'Yes' to one of the two questions, and 'No' to the other one 2. 'Yes' to both questions, or 'No' to both questions	C28