// NO.19-055 | 12/2019

DISCUSSION PAPER

// ARMENAK ANTINYAN AND
ZAREH ASATRYAN

Nudging for Tax Compliance: A Meta-Analysis

Nudging for Tax Compliance

A Meta-Analysis

Armenak Antinyan*† Zareh Asatryan*‡

November 15, 2019

Abstract

Taxpayer nudges – behavioral interventions that aim to increase tax compliance without changing the underlying economic incentives of taxpayers – are used increasingly by governments because of their potential cost-effectiveness in raising tax revenue. We collect about a thousand treatment effect estimates from over 40 randomized controlled trials, and in a meta-analytical framework show that non-deterrence nudges – interventions pointing to elements of individual tax morale – are on average ineffective in curbing tax evasion, while deterrence nudges – interventions emphasizing traditional determinants of compliance such as audit probabilities and penalty rates – are potent catalysts of compliance. These effects are, however, fairly small in magnitude. Deterrence nudges increase the probability of compliance by only 1.5-2.5 percentage points more than non-deterrence nudges, while the effects are likely to be bound to the short-run, and are somewhat inflated by selective reporting of results.

JEL codes: C93, D91, H26

Keywords: Tax compliance, Randomized control trials, Nudging, Meta-analysis.

*We thank Sebastian Blesse, Annika Havlik, Jost Heckemeyer, Friedrich Heinemann, Carla Krolage, Carina Neisser, Anh Pham, Johannes Rincke as well as workshop participants at ZEW Public Finance Conference, Armenian Economic Association and International Institute of Public Finance for valuable comments. We are grateful to Felix Köhler, Agon Topxhiu, David Westerheide and Zeyuan Xiong for excellent research assistance.
†Zhongnan University of Economics and Law, antinyan.armenak@gmail.com.
‡ZEW Mannheim, zareh.asatryan@zew.de.

1 Introduction

Recent years have seen a lot of excitement around the idea of using "nudges" with the aim to improve individual behavior. Nudges are interventions that respect the freedom of choice and leave economic incentives intact (Benartzi et al. 2017),[1] and they have been studied in many important policy areas such as taxation (Mascagni 2018), education (Dizon-Ross 2019), healthcare (Wisdom et al. 2010), consumer behavior (Costa and Kahn 2013), among others.

In the field of taxation,[2] nudging has become widely popular in the last decade among policy makers who often claim that relative to the negligible direct cost of nudging (e.g., sending a letter) the potential payoffs involved can be extremely high.[3] Academic economists, on the other hand, have come to recognize that a large behavioral response to a simple informational update induced by a nudge is not overly consistent with expected utility theories of human behavior where tax compliance is primarily driven by fundamental economic incentives.[4] Such evidence showing that agents respond to nudges would at least demonstrate the presence of information imperfections (e.g., taxpayers misperceive the probability of being detected when evading), and would possibly hint to the existence of deviations from utility maximization (e.g., taxpayers additionally care about tax morale).[5]

[1] Thaler and Sunstein (2008) define a nudge as an "aspect of the choices architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives." They continue that for an intervention "to count as a mere nudge, the intervention must be easy and cheap to avoid".

[2] The most widely used nudge intervention in the field of taxation unfolds as a structured communication campaign where the tax administration delivers some content to the taxpayers that qualifies to the properties of being a nudge as defined above by Thaler and Sunstein (2008). Sunstein (2014) presents a list of common nudges.

[3] For instance, Hallsworth et al. (2017) and Bott et al. (2017) report £9 million and $25 million increase in tax revenues, respectively, due to letters sent. Although, typically these letters are interpreted of being virtually costless, Allcott and Kessler (2019) argues that nudges entail significant costs for the nudge recipients and shows that the failure to take into account these costs overstates the effects of nudges on social welfare.

[4] This canonical theory of Allingham and Sandmo (1972), Yitzhaki (1974), modeled after the economics of crime literature (Becker 1968), views the individual as a rational agent with some level of risk aversion who considers tax evasion as a gamble trading off the benefits of successful evasion against the costs of detection and punishment.

[5] Motivated by mounting evidence on various behavioral biases of taxpayers, Farhi and Gabaix (2019) develops a theory of optimal taxation with behavioral agents also incorporating nudges into this framework. For a recent review of behavioral public economics literature, see, Bernheim and Taubinsky (2018).

In this paper, we use methods of meta-analysis[6] and ask whether nudges are really effective in increasing tax compliance levels among individuals and small firms. In so doing we aim to present a systematic review of the literature and to provide guidance for further tax experiments and policy interventions. Of course, an alternative and arguably a more thorough way of surveying the literature can be done through qualitative means. Many excellent literature reviews have been written, such as Alm (2019), Andreoni et al. (1998), Slemrod (2007, 2019), Slemrod and Yitzhaki (2002) on tax compliance generally, as well as Luttmer and Singhal (2014) on tax morale, Mascagni (2018) on tax experiments and Pomeranz and Vila-Belda (2019) on tax capacity, more specifically. We extend this literature by performing a systematic empirical analysis of nudging interventions. This is an important task since as Luttmer and Singhal (2014) put it "similar interventions have produced varying results in different contexts" and "it would be useful to examine why". We are aware of two meta-studies of tax experiments (Alm et al. 2018, Blackwell 2007) which study laboratory experiments while we focus on field work.[7]

As opposed to qualitative reviews, our meta-analysis can give more systematic answers to questions like: i) Are nudges effective in curbing tax evasion? ii) If so, on which margins of compliance and by how much on average? iii) Which nudge types work more effectively? iv) Are nudges effective also in the longer horizon? v) Which groups of taxpayers are more responsive to nudges?

To answer these questions we collect data on 940 treatment effect estimates of tax compliance coming from 41 studies.[8] We divide these data into three different samples according to the measure of tax compliance employed.[9] First is the full sample of t-values

[6]For a review of these methods, see, Anderson and Kichkha (2017), Nelson and Kennedy (2009), Stanley et al. (2013), Stanley (2001), Stanley and Doucouliagos (2012).

[7]Blackwell (2007) concludes that increasing the penalty rate, the marginal per capita return to the public good and the probability of audit lead to higher tax compliance, meanwhile tax rate has no significant impact on tax compliance. Focusing on a larger set of papers, Alm et al. (2018) illustrate that audit probability increases tax compliance on the extensive margin, while audit probability and the tax rate influence tax compliance negatively on the intensive margin.

[8]These studies are listed in Table 1. The map in Figure 1 shows the geographic distribution of these experiments as well as of 18 ongoing interventions registered in the randomized control trial (RCT) registry of the American Economic Association (as retrieved on October 21, 2019).

[9]Figures 2, 3(a) and 3(b) present the distributions of these three variables.

where we can study the direction and statistical significance of treatment effects. Second sub-sample includes 440 observations from 27 papers that study tax compliance at the extensive margin measured as either the probability to pay or file taxes. Third sub-sample comprises of 172 observations from 17 papers that study tax compliance at the intensive margin measured as the (log) amount of taxes paid.

Unlike many other meta-studies in economics,[10] one advantage of this paper is that we pool together RCTs which have a relatively high degree of homogeneity in their quality of identification. The high level of comparability of treatment effects coming from different interventions makes our conclusions more meaningful.[11] Another favorable feature of this exercise is that we can study magnitudes of effects in addition to their direction since tax compliance, our dependent variable of interest, can be measured in a relatively standard form. A third advantage is that since almost all of the studies in our sample implement several interventions in their experiments, we can use study fixed effects thus obtaining within study estimates of nudges that control for all study characteristics.

Our main finding is that, in contrast to the recent excitement over nudges, the behavioral content introduced in the communication between the tax administration and the taxpayer is not as effective as often thought. We present robust evidence that on average only deterrence interventions, i.e., nudges informing about audit probabilities and potential penalties, work in increasing compliance levels. The effects of behavioral letters that inform taxpayers about the importance of paying taxes for the adequate provision of public goods, about the (positive) behavior of their peers, or hint towards general appeals of paying taxes as a moral obligation are on average ineffective. The effect magnitudes of deterrence intervention are moderate as they increase compliance by 1.5 to 2.5 percentage points on the extensive margin, and 5 to 15 percentage points on the intensive margin compared to taxpayers receiving non-deterrence

[10] For several recent applications, see, Card et al. (2010, 2017), Feld and Heckemeyer (2011), Gechert (2015), Heinemann et al. (2018), Lichter et al. (2015), Neisser (2017).

[11] This argument is one reason behind the methods of meta-analysis being so much more popular in the field of medical sciences (which often evaluate randomized clinical trials) than in economics (Stanley 2001).

treatments. These effects are also likely to be bound to a short-run, and may be somewhat inflated by selective reporting of results.[12]

A strong way of interpreting this evidence is that individual financial motives, rather than elements of tax morale[13] like social norms or reciprocity remain the first order factors behind compliance decisions. A competing interpretation, however, is that deterrence and non-deterrence nudges are not equally effective in shifting the prior beliefs of taxpayers. According to Pomeranz and Vila-Belda (2019), it may well be that nudges implemented by tax authorities are more effective in shifting perceptions of audit probabilities than perceptions of social norms. Either way, our evidence suggests that at the very least the mainstream neoclassical approach to tax evasion should take into account the possibility that taxpayers are constrained with information imperfections.

Our additional findings highlight certain design aspects of experiments that make them more effective. For example we find that nudges communicated through in-person visits deliver more powerful results in terms of compliance outcomes than nudges communicated through letters. In terms of different groups of taxpayers, we find that groups having more non-compliers (such as late-payers) are more likely to be affected by nudges, while groups of taxpayers typically being less susceptible to non-compliance (such as VAT taxpayers in accordance with arguments of third-party reporting and paper-trail) are less sensitive to nudges. In general, the types of nudges explain about 34% and 23% of the observed within study variation in treatment effect estimates of, respectively, extensive and intensive margin responses. Meanwhile, the additional characteristics of the studies capture another 13% to 19% of the remaining heterogeneity.

[12]Both in terms of p-hacking, where marginally significant treatment effects are more likely to be reported than results narrowly failing to reject the null, as well as file drawer type of bias, where the results not supporting the likely hypotheses of researchers are not reported.

[13]See Besley et al. (2019) for theory and evidence on the interaction between individual and social motives in tax evasion.

2 Selection of studies

Literature search: We ran a literature search on a rolling basis throughout March to October of 2019. First, we searched for relevant papers using a defined combinations of keywords[14] in the main literature databases of the profession.[15] Second, to identify ongoing work, we continued the search in the programs of the main general interest conferences of in economics as well as the main conferences specializing on behavioral or experimental economics and public economics.[16] Third, we carefully looked through the bibliographic information of the papers identified in the last two steps to further refine the study sample.

Table 1: List of studies in meta-analysis sample

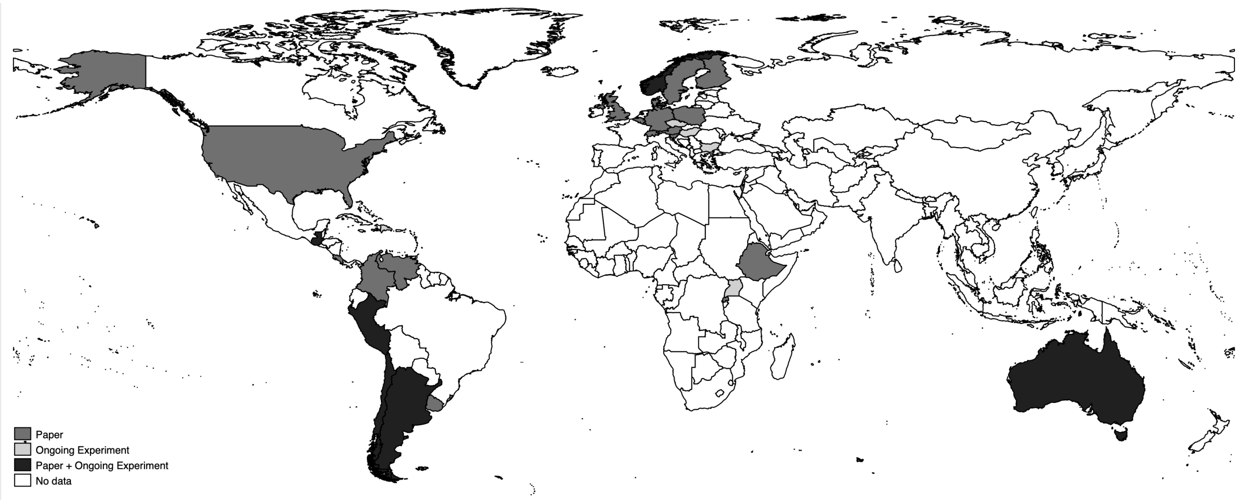
No	Author	Country	No	Author	Country
1	Appelgren (2008)	Sweden	22	Hasseldine et al. (2007)	UK
2	Ariel (2012)	Israel	23	Hernandez et al. (2017)	Poland
3	Bérgolo et al. (2017)	Uruguay	24	Hiscox (Hiscox)	Australia
4	Biddle et al. (2018)	Australia	25	Iyer et al. (2010)	USA
5	Blumenthal et al. (2001)	USA	26	John and Blume (2018)	UK
6	Boning et al. (2018)	USA	27	Kettle et al. (2016)	Guatemala
7	Bott et al. (2017)	Norway	28	Kettle et al. (2017)	Guatemala
8	Boyer et al. (2016)	Germany	29	Kleven et al. (2011)	Denmark
9	Brockmeyer et al. (2019)	CostaRica	30	Mascagni et al. (2017)	Rwanda
10	Castro and Scartascini (2015)	Argentina	31	Mascagni et al. (2018)	Ethiopia
11	Chirico et al. (2019)	USA	32	Meiselman (2018)	USA
12	Coleman (1996)	USA	33	Ortega and Sanguinetti (2013)	Venezuela
13	Del Carpio (2013)	Peru	34	Ortega and Scartascini (2015)	Columbia
14	De Neve et al. (2019)	Belgium	35	Perez-Truglia and Troiano (2018)	USA
15	Doerrenberg and Schmitz (2017)	Slovenia	36	Pomeranz (2015)	Chile
16	Dwenger et al. (2016)	Germany	37	Scartascini and Castro (2019)	Argentina
17	Eerola et al. (2019)	Finland	38	Shimeles et al. (2017)	Ethiopia
18	Fellner et al. (2013)	Austria	39	Slemrod et al. (2001)	USA
19	Gillitzer and Sinning (2018)	Australia	40	Torgler (2004)	Switzerland
20	Hallsworth et al. (2017)	UK	41	Wenzel (2006)	Australia
21	Harju et al. (2018)	Finland			

[14]The keywords include: randomized controlled trial, RCT, field experiment, nudging, nudges, behavioral intervention, tax evasion, tax compliance, tax non-compliance.

[15]The literature databases include: Econlit, Google Scholar, and Science Direct.

[16]The conferences include: American Economic Association, European Economic Association, ESA, SABE, WEAI, National Tax Association, International Institute of Public Finance.

Figure 1: Country coverage of nudging experiments



Source : Own compilation. Data on ongoing experiments is based on the RCT registry of the American Economic Association.

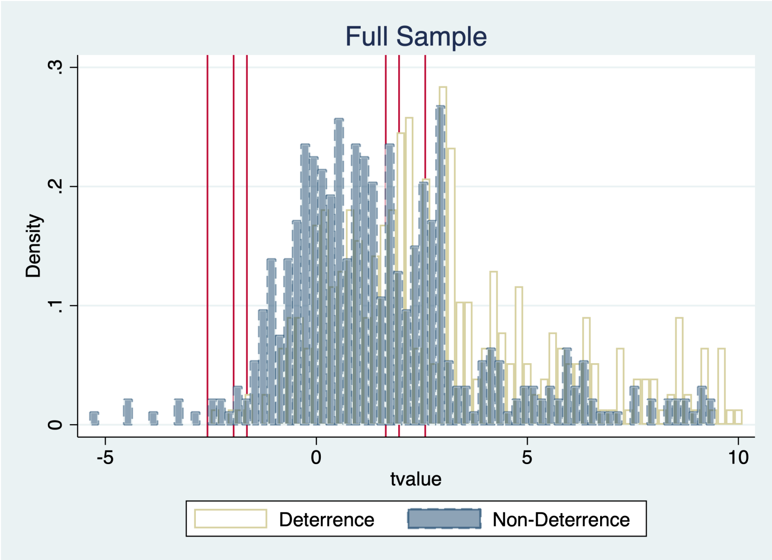
Study inclusion criteria: For a paper to be included in our sample all of the following four criteria need to be fulfilled: i) the study is based on a RCT performed at the level of taxpayers (i.e. individuals or firms rather than, e.g., regions); ii) the trial introduces a nudging intervention which closely follows the definition of Thaler and Sunstein (2008); iii) the dependent variable of interest is the tax payment behavior of the taxpayer; and iv) the resulting study reports all the relevant statistics necessary for our meta-analysis (e.g., effect sizes along with the standard errors) for at least one treatment effect estimate.

Final sample: After applying these four filters to the list of papers collected from our extensive search we arrive at an overall sample of 41 studies. These studies are listed in Table 1 in alphabetical order. These 41 experiments were performed in 24 countries situated mainly in Europe, Africa, Australia and the Americas as presented in the map of Figure 1.

3 Tax compliance measures and types of nudges

Dependent variables: Our total sample comprises of 940 estimates collected from 41 papers. From these data we define three different dependent variables. First is the full sample

Figure 2: Distribution of t-values in the full sample



Notes : Figure plots a histogram of t-values of treatment effects which we obtained from the primary studies in our sample. We plot t-values for deterrence and non-deterrence nudges separately. For visual clarity, we drop 104 outlier observations that lie outside the (-10, 10) range. Vertical lines denote critical values for two-sided significance tests at t-values of +/-1.645, +/-1.96 and +/-2.58.

of t-values where we can study the direction and statistical significance of effects. Figure 2 presents a histogram of the distribution of t-values. We transform this variable into a dummy variable that equals 1 when t-values are larger than 1.96 (i.e. the critical value of rejecting the null hypothesis at 5% level of statistical significance with a two-sided test) and 0 otherwise.[17]

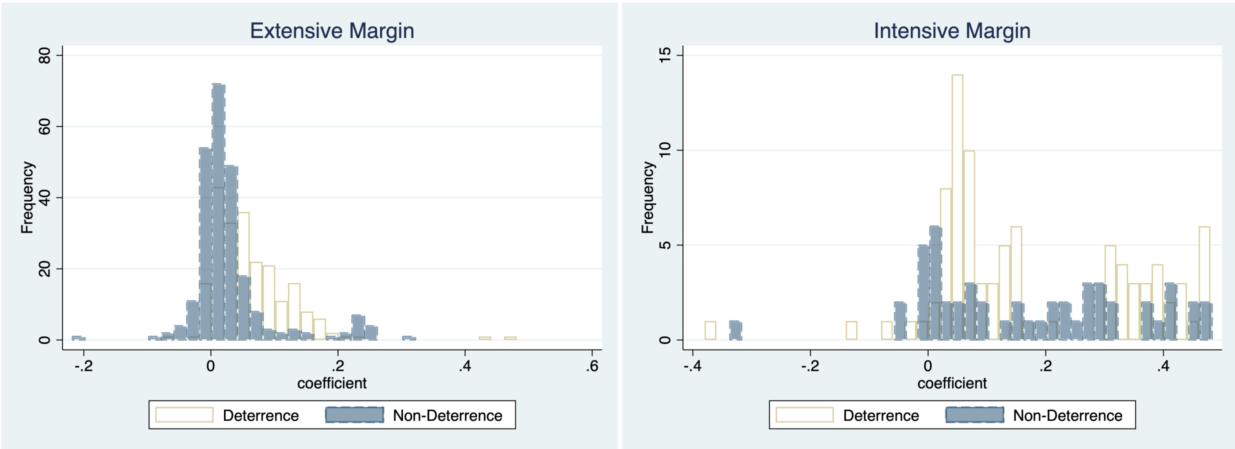
Our second and third dependent variables measure the magnitude of tax compliance. In particular, we differentiate between treatment effect estimates that measure the impact of nudges on extensive margin of tax compliance (i.e., the probability of compliance), and estimates that measure the impact of nudges on intensive margin of tax compliance (i.e., the extent of compliance).[18] Since the studies under consideration measure tax compliance in

[17] Such a transformation is common in the meta-analytical literature (see, e.g., Card et al. 2017). We pool and code the negative and significant treatment effects together with insignificant effects rather than coding them separately, since as shown in Figure 2 we have very few negative and significant effects. In our baseline results we define statistical significance at the 5% cutoff, but the results remain stable if we define them at 1% or 10% levels.

[18] We do not claim that the intensive margin effect is necessarily separable from the extensive margin effect in all contexts, and rather rely on the respective specification of the primary study. For work that allows for both intensive and extensive margin responses to taxes, see, Blundell and MaCurdy (1999) and Kleven and Kreiner (2006) in the context of labor supply responses, and Almunia et al. (2019) in the context of tax deductible charitable donations.

Figure 3: Distribution of treatment effects of extensive and intensive margin responses

(a) Probability to comply **(b)** Log amount of tax paid



Notes : Sub-figure (a) and (b) plot a histograms of treatment effects on extensive and intensive margins of compliance, respectively, which we obtained from the primary studies in our sample. We plot these treatment effect for deterrence and non-deterrence nudges separately. For visual clarity, sub-figures (a) and (b) drop, respectively, 8 and 30 outlier observations that are larger than 0.5.

various ways we lose a substantial number of observations, when we restrict our dependent variable to either the extensive or the intensive margin of tax compliance. More specifically, our extensive margin variable includes 440 observations from 27 studies and is measured as either the probability to pay or file taxes or report taxes. The third dependent variable comprises of 172 observations from 17 papers that study tax compliance at the intensive margin measured as the log amount of taxes paid. Figures 3 (a) and (b) present histograms of the distribution of treatment effects on extensive and intensive margins of compliance, respectively.

Deterrence nudges: We classify nudges into two main deterrence and non-deterrence categories. Each category contains about half of our full sample. To be considered as a deterrence nudge, the communication between tax administration and taxpayers should specifically contain a threat that highlights one of the economic factors behind the tax compliance decision as in the canonical model of tax evasion by Allingham and Sandmo (1972): mainly the possibility of audit and the potential penalty if caught evading. An example of such a nudge is the following one used by (Castro and Scartascini 2015): "Did you know that if you do not pay the CVP on time for a debt of AR$ 1,000 you will have to disburse AR$ 268 in arrears at the end of the year and the Municipality can take administrative and legal action?".[19]

Main non-deterrence nudges: To be acknowledged as a non-deterrence nudge, the communication content between the tax administration and taxpayers should not contain a threat that has the potential to alter the taxpayers' financial motives. The first and main sub-category of non-deterrence nudges is built on the solid evidence in the literature that taxpayers can be motivated by such considerations as morality, the perception of fairness, social norms in the society, provision of public goods by the government and the like. The common nudges in this sub-category are detailed below:

- Public good, which makes it clear that the taxes paid by individuals are used to finance public goods and services: "Your tax payment contributes to the funding of publicly financed services in education, health and other important sectors of society" (Bott et al. 2017).
- Peer effect, which underlines that the majority of individuals in a given country/community are complying with taxes: "Nine out of ten people pay their taxes on time" (Hallsworth et al. 2017).
- Moral appeal, which appeals to morality to influence taxpayer behavior: "If the taxpayers did not contribute their share, our commune with its 6226 inhabitants would suffer greatly. With your taxes you help keep Trimbach attractive for its inhabitants" (Torgler 2004).

In our analysis we study deterrence and the main non-deterrence nudges as two coherent and broad groups of nudge types, and in additional analysis we study the individual effects of these three types of non-deterrence nudges separately.

[19]Note that communications including both deterrence and non-deterrence components are classified as a deterrence nudge given the presence of the threat component. In an additional analysis we label these communications as mixed deterrence nudges and study their effects separately.

Other non-deterrence nudges: The second sub-category of non-deterrence nudges mainly contains manipulations that are utilized "to correct" the taxpayer non-compliance that stems from such behavioral fallacies as limited attention, procrastination and cognitive overload among others. For instance, simple reminders are sent to taxpayers to overcome the problem of limited attention (Hernandez et al. 2017, Mascagni et al. 2017). The problem of cognitive overload is usually bypassed through the simplification of the communication language, the introduction of visual stimuli or provision of information how to file the income (De Neve et al. 2019, Eerola et al. 2019). This sub-category of non-deterrence nudges also include communications which introduce various types of informational content, such as a sentence on tax deductible donations (Biddle et al. 2018), phone number for enhanced consumer service (Coleman 1996), a statement of intent by tax administration to help during the filing process (Hasseldine et al. 2007), and the like. In the analysis that follows, we group all these nudges under the common umbrella name Other Non-Deterrence nudges. Although, the types of nudges in these sub-category are not always coherent in the type of content they introduce, they make only about a quarter of non-deterrence nudges.

Study characteristics: We account for a number of characteristics that vary within as well as between studies. In particular we study: i) type of the tax, i.e., personal income tax, corporate income tax, property tax, VAT, all taxes and other taxes[20]; ii) communication channel used by the tax authority to reach out to the taxpayers, i.e., digital letters (e.g., e-mails, SMS, CAPTCHA), physical letters (e.g., letters, tax bill manipulations), and in-person visits; iii) response horizon of the compliance measure, i.e., a dummy variable on whether the time interval between the date on which the nudge was sent and the date when the outcome variable was measured is shorter or longer than 12 months; iv) the benchmark against which the interventions are evaluated, i.e., the control group did not receive any communication (no information), received some neutral information (baseline information, or the comparison is made against another behavioral intervention (letter)); v) the estimation method of the

[20]Other taxes include country-specific taxes or fees, e.g., the church tax in Germany, wealth tax in Colombia, TV license fees in Austria.

experimental data, i.e.,difference-in-difference, OLS, non-linear, etc.; vi) a dummy whether the taxpayer is a late payer, i.e., did not comply in paying taxes by the official deadline; vii) type of the taxpayer, i.e., individuals, business entities, or a sample including both types; viii) the number of observations in the experiment; ix) publication status of the study, i.e., working paper or published article; and x) the country where the experiment was conducted.

4 Empirical methodology

Our main interest is to study the effects of deterrence nudges relative to non-deterrence types of nudges on tax compliance.[21] Almost all of the studies in our sample implement several interventions.[22] Therefore, unlike many other meta-analyses in economics, we can exploit the substantial within-study variation in the data and control for study fixed effects. In extended results, we are also interested in the question of whether various study-, experiment-, or country-specific characteristics drive the heterogeneity in results. Some of these characteristics do not vary within studies, therefore in additional specifications we drop the study fixed effects and compare the average effects of characteristics across studies.

We estimate the following equation:

$$Estimate_{i,p}^{\tau} = \alpha + \beta^{\tau} Nudge_{i,p} + \gamma Controls_{i,p} + \lambda_p + \epsilon_{i,p} \tag{1}$$

where $Estimate_{i,p}^{\tau}$ is the i^{th} estimate from paper p of type τ: i) transformed t-value of primary studies into a dummy variable that equals 1 for treatment effect estimates larger than 1.96, i.e., that are positive and statistically significant at least at the 5% level, and 0 otherwise; ii) treatment effects on extensive margin responses, and iii) treatment effects on intensive margin

[21]The total sample consists of about 940 estimates collected from around 41 primary studies. Around half of these treatment effects are obtained from deterrence interventions and the other half come from non-deterrence interventions.

[22]Since the number of estimates across studies differs, we treat our data as an unbalanced panel. Two studies do not allow for within-study variation, that is, Blumenthal et al. (2001) with only one estimate and Torgler (2004) with one type of nudging intervention (moral suasion).

responses.[23] $Nudge_{i,p}$ is a binary variable which equals 1 in case of a deterrence intervention and 0 in case of a non-deterrence intervention. In additional specifications we define $Nudge_{i,p}$ more broadly as a categorical variable taking into account the several different types of nudges as defined in Section 3. λ_p captures the study fixed effects. $\epsilon_{i,p}$ is the error term which is clustered at the level of papers p.[24]

β is our main coefficient of interest which shows the relative effect of different types of nudges on the tax compliance measure under study. However, we are also interested in whether the control variables in $Controls_{i,p}$ can explain the heterogeneity in the tax compliance measure. Thus, in Equation 1, we introduce the variables i) to x) discussed in Section 3 either one by one or jointly. Table A1 of the Appendix shows the summary statistics of all dependent and independent variables.

In the choice of our estimation methods we follow a number of recent applications of meta-analytical techniques in economics (Card et al. 2010, 2017, Feld and Heckemeyer 2011, Gechert 2015, Heinemann et al. 2018, Lichter et al. 2015, Neisser 2017) as well as a literature reviewing these methods (Nelson and Kennedy 2009, Stanley et al. 2013, Stanley 2001, Stanley and Doucouliagos 2012). Our simplest specification relies on an OLS estimator and includes paper fixed effects. Meta-analytical regressions are known to be heteroskedastic,[25] therefore we follow the literature and as a second specification use a WLS estimator, where as analytical weights we take the inverse of the squared standard error of the parameter estimates.[26] To account for the unbalanced nature of our panel data, in a third specification we follow Heinemann et al. (2018) and replace the former analytical weights with inverse of the share of observations per study in relation to the full sample. Note that this latter method of weighting is also applicable to our full sample of t-values where we study the statistical significance and direction of estimates. While the former weighting scheme that

[23] Figures 2, and 3(a) and 3(b) plot the distributions of the three dependent variables, respectively. For a more detailed description of these variables, see Section 3.

[24] Estimates are obvioulsy not independent within studies, therefore we choose to cluster errors at this level.

[25] One form of heteroskedasticity arises because the variance of the individual estimates is negatively related to the size of the underlying sample and this correlation is likely to be different between the primary studies.

[26] Due to their wide distribution we follow Card et al. (2017) and winsorize these analytical weights at the top and bottom deciles. Results remain very similar to alternative winsorizations at 1 or 5 percentiles.

makes use of standard errors is more standard in the literature, it is obviously appropriate to use only in specifications where we study the magnitudes of compliance at the extensive and intensive margins. As a fourth and final estimator we adopt a random effects model,[27] which assumes the existence of a distribution of true effects for distinct studies and populations. Thus we relax the assumption that for each type of a nudge there exists a single "true" effect which is common to all studies under consideration.

5 Results

We start by describing our baseline results in Sub-section 5.1, which compare the relative effects of deterrence nudges to those of non-deterrence nudges within papers. In the following Sub-section 5.2 we study the relative effects of the more detailed categories of nudges as well as extend the discussion from relative effects to average effects. Sub-section 5.3 relaxes our baseline identification strategy of using study fixed effects, which allows to study the role of a wider set of study characteristics in addition to types of nudges in explaining the variation in estimated treatment effects. Sub-section 5.4 adopts a number of approaches to examine traces of publication bias in our sample of RCTs.

5.1 Baseline results: Deterrence and non-deterrence nudges

Table 2 studies the effect of deterrence nudges relative to non-deterrence nudges on the statistical significance and magnitudes of treatment effect estimates. The dependent variable in Panel A is the t-value of primary studies transformed to a dummy variable that equals 1 for estimates that are positive and statistically significant at least at the 5% level, and 0 otherwise. The dependent variable in Table 2 Panel B measures tax compliance at the extensive margin, while the one in Panel C measures compliance at the intensive margin.

All regressions include study fixed effects. Odd numbered columns represent the baseline model, while even numbered columns include the full set of control variables (not reported

[27]The terminology of random effects in this context should not be confused with the study fixed effects, the inclusion of dummies for individual studies.

Table 2: Baseline Results

	(1) WLS(se)	(2) WLS(se)	(3) OLS	(4) OLS	(5) WLS(n)	(6) WLS(n)	(7) RE	(8) RE
PANEL A: FULL SAMPLE								
Deterrence			0.203^{*}	0.188^{*}	0.246^{**}	0.221^{*}	0.221^{***}	0.212^{***}
			(0.0855)	(0.0879)	(0.0796)	(0.0850)	(0.0351)	(0.0348)
Observations			940	924	932	916	924	908
PANEL B: EXTENSIVE MARGIN								
Deterrence	0.0137^{**}	0.0137^{**}	0.0245^{***}	0.0247^{***}	0.0208^{**}	0.0213^{**}	0.0253^{*}	0.0254^{**}
	(0.00414)	(0.00443)	(0.00577)	(0.00586)	(0.00695)	(0.00722)	(0.00986)	(0.00899)
Observations	440	436	456	452	456	452	440	436
PANEL C: INTENSIVE MARGIN								
Deterrence	0.0503^{*}	0.0617^{*}	0.126^{*}	0.137^{*}	0.0901^{*}	0.0980	0.119	0.128
	(0.0194)	(0.0236)	(0.0449)	(0.0497)	(0.0332)	(0.0467)	(0.323)	(0.310)
Observations	172	172	172	172	172	172	172	172
Study fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	No	Yes	No	Yes	No	Yes	No	Yes

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes : Regressions are estimated according to Equation 1. The dependent variables are: i) Panel A: the t-value of primary studies transformed to a dummy variable that equals 1 for estimates larger than 1.96 (i.e., that are positive and statistically significant at least at the 5% level) and 0 otherwise; ii) Panel B: the treatment effect on extensive margin of tax compliance (probability of compliance); iii) Panel C: the treatment effect on intensive margin of tax compliance (log amoun of taxes paid). The independent variable of interest is a dummy equal to 1 for all deterrence treatments, and 0 for all non-deterrence treatments (ommited category). WLS(se) and WLS(n) are weighted least squares estimators which as analytical weights take, respectively, the inverse of the squared standard error of the parameter estimates (winsorized at top and bottom deciles) and the inverse of the share of observations per study in relation to the full sample.

in the tables). The table presents these two specifications with and without control variables across the four alternative estimators as introduced in Section 4.

The most basic OLS estimators collected in Table 2 Panel A show that the studies in our sample are 20% more likely to find that deterrence nudges increase tax compliance compared to the effects of non-deterrence nudges (column 3). This result is robust to the inclusion of our full set of control variables (column 4). The magnitude of this effect somewhat increases to 22-25% when taking into account the share of estimates per primary study by the WLS estimator (columns 5-6). Random effects estimator yields results that are similar in magnitude and are more precisely estimated (columns 7-8).[28]

[28]Since the dependent variable in Table 2 Panel A is based on a transformation of t-values, we can not implement a WLS estimator using the standard errors as analytical weights. The estimations using this method are otherwise presented in columns 1-2.

Our sample size decreases by around half when studying the sub-sample of estimates that measure tax evasion at the extensive margin, but we still have sufficient variation coming from around 27 studies for identification. The dependent variable is measured homogenously across the papers and is defined as the probability of a taxpayer to comply with her taxes in a given time horizon. Table 2 Panel B collects the estimates. Consistent with the previous results we find that deterrence nudges are more effective in improving the level of tax compliance than non-deterrence nudges. The effects are robust across all the specifications. The magnitudes of these effects are fairly small, however, and are in the range of 1.4 to 2.5 percentage point increase in the probability to comply when receiving deterrence as opposed to receiving non-deterrence nudges.

Table 2 Panel C studies tax compliance at the intensive margin. Studies use different measures of intensive margin responses and the most common one of these that we focus on here studies the log amount of taxes paid. We are left with only about the sixth of our original sample size which substantially reduces the precision of our estimates. Across the eight specifications of Table 2, five are statistically significant at the 10% level. Nevertheless, these results are consistent with the above findings and they indicate that deterrence nudges increase compliance levels by 5 to 13% more than the non-deterrence nudges (depending on the estimator under consideration).

5.2 Types of nudges and the baseline effect

In this section we extend the baseline analysis as presented in Section 5.1 in two ways. First, instead of grouping the multiple types of behavioral interventions into deterrence and non-deterrence nudges, we more flexibly study whether each of these types of interventions have effects on tax compliance. More specifically, within the non-deterrence nudge category we distinguish between the main non-deterrence categories we are interested in – that are Public Goods, Peer Effects, and Moral Appeals – and Other Non-Deterrence nudges. Within the deterrence nudge category we also differentiate between pure deterrence and mixed deterrence nudges, where the latter type of nudges append a deterrence with a non-deterrence

Table 3: All Treatments

	(1) Full	(2) Full	(3) Extensive	(4) Extensive	(5) Intensive	(6) Intensive
Omitted: Other Non-Deterrence nudges						
Deterrence	0.178	0.158	0.0212*	0.0227*	0.0606	0.0641
	(0.0903)	(0.0907)	(0.00880)	(0.00951)	(0.105)	(0.124)
Deterrence Mixed	0.382**	0.346**	0.0276**	0.0249***	0.177**	0.168*
	(0.117)	(0.127)	(0.00965)	(0.00665)	(0.0502)	(0.0593)
Public Good	-0.0187	-0.0288	-0.00653	-0.00572	-0.0780	-0.0946
	(0.0707)	(0.0719)	(0.00830)	(0.00868)	(0.105)	(0.108)
Peer Effect	0.0467	0.0355	0.00230	0.00353	-0.0863	-0.0930
	(0.113)	(0.116)	(0.0121)	(0.0123)	(0.0879)	(0.0940)
Moral Appeal	0.0204	0.00683	-0.000667	0.000513	-0.0315	-0.0286
	(0.108)	(0.109)	(0.00855)	(0.00902)	(0.0910)	(0.106)
Constant	-0.178	-1.195***	0.0497***	0.173	0.0771	-0.0667
	(0.0903)	(0.234)	(0.00545)	(0.111)	(0.102)	(0.626)
Study fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Control variables	No	Yes	No	Yes	No	Yes
Observations	940	924	456	452	172	172
R^2	0.437	0.496	0.340	0.465	0.234	0.419

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes : Regressions are estimated according to Equation 1. The dependent variables are: i) Columns 1-2: the t-value of primary studies transformed to a dummy variable that equals 1 for estimates larger than 1.96 (i.e., that are positive and statistically significant at least at the 5% level) and 0 otherwise; ii) Columns 3-4: the treatment effect on extensive margin of tax compliance (probability of compliance); iii) Columns 5-6: the treatment effect on intensive margin of tax compliance (log amoun of taxes paid). The independent variable of interest is a categorical variable capturing deterrence treatments, deterrence mixed treatments, several types of non-deterrence treatments and a group of other non-deterrence treatments (ommited category).

communication. Second, while in Section 5.1 we showed that deterrence nudges are effective in increasing compliance as compared to non-deterrence, we are additionally interested in the question of whether these nudges are effective relative to some baseline. For that reason, we omit the category of Other Non-Deterrence nudges and study the relative effects of deterrence and non-deterrence nudges to this reference category.[29]

The estimates are collected in Table 3 for the three dependent variables of interest. As in Table 2 all regressions include study fixed effects. To preserve space we report results only for the simple OLS estimator rather than across all four estimators as in Table 2.[30]

[29] As discussed in Section 3, the category of other non-deterrence nudges include reminders to pay taxes, information on how to file and the like. We believe that, at least in theory, such interventions should not facilitate responses that work similiar to the underlying response mechanism of the deterrence and main non-deterrence nudges. Therefore, this category of other non-deterrence nudges may serve as a reference point to study the effect of deterrence nudges and the main non-deterrence interventions aimed at influencing taxpayers' tax morale.

[30] The results are robust to these estimators and are available upon request.

Once we distinguish between deterrence and mixed deterrence nudges, i.e. deterrence nudges that additionally include one of the main types of non-deterrence nudges, the results on the deterrence category of our treatment variable become somewhat less precise although the magnitudes are robust to what we found in the baseline results. Although the mixed deterrence nudges become strong and sizable, in an additional test where we omitted the simple deterrence category we failed to reject the null hypothesis that mixed deterrence nudges are not different from deterrence nudges. Regarding the individual effects of three main types of non-deterrence nudges – Public Goods, Peer Effects, and Moral Appeals – we do not find evidence that they are statistically distinguishable from zero.

Importantly, these results that compare the types of nudges we are interested in to the ommited category of neutral nudges support our claims that in addition to the relative effects of deterrence and non-deterrence nudges, our evidence supports the more general conclusion that deterrence nudges are effective in increasing compliance while non-deterrence ones are not. This interpretation is reinforced if we drop study fixed effects and look at the baseline effect of nudges. The point estimate on the constant term in the regression without any control variable and nor study fixed effects (column 2 of Table 4) shows that in the baseline about half of treatment effects find positive and significant effects. Once we control for the type of the nudge (column 3) the point estimate on the constant term decreases by about 30%, while the inclusion of other study characteristics in addition to the nudge type (column 10) fully explains the remainder of the variation in the baseline effect forcing it to nearly reach zero.

5.3 Explaining the heterogeneity in treatment effects

The last two sections have documented our baseline finding that deterrence interventions are effective in increasing tax compliance while non-deterrence interventions are not. By accounting for the types of nudges we were able to explain about 34% and 23% of the observed within study variation in treatment effect estimates of, respectively, extensive and intensive margin responses (see R^2 in columns 3 and 5 of Table 3). In this section we study how the various additional study characteristics drive our results. Taken together these study characteristics can explain an additional of 13% to 19% of the remaining heterogeneity in the treatment effect estimates of nudges (see the difference in R^2 between Table 3 columns 3 and 4, and 5 and 6).

Some of these characteristics vary within studies. For example, several papers estimate tax compliance responses across different time horizons which allows to study whether the effect of nudges diminishes over time controlling for study fixed effects. However, more often such characteristics do not vary within studies. For example, we do not have a single experiment in our sample that was implemented across multiple countries. Therefore, we relax our baseline model here and include study fixed effects only in one specification as a robustness test. Otherwise, we test whether our control variables systematically correlate with the tax compliance measure by including them in the specification first one by one and then jointly. The variation in most study characteristics we have is limited even after dropping the study fixed effects. Therefore, in this section we primarily study the full sample of t-values as shown in Table 4. Estimations on the sub-samples capturing extensive and intensive margin responses, which contain, respectively, around half and one-sixth of the full sample, are reported in Table A2 of the Appendix. Table A1 of the Appendix shows the summary statistics of these chacateristics in each of the three sub-samples.

Several results stand out. First, Table 4 shows evidence for the hypothesis that the treatment effects are stronger statistically in the short-run compared to the long-run. In particular, column 10 shows that treatment effects where compliance is measured within 12 month after the interventions are 26% more likely to be statistically significant than treatment effects measuring compliance after 12 months of the intervention. This result holds even when exploiting within study variation. When restricting the analysis to the extensive and intensive margin sub-samples in Table A2 we do not find evidence that the magnitude of compliance is different in short compared to long horizons. One explanation is that even if nudge interventions are more likely to yield to significant treatment effects in the short run,

Table 4: Heterogenous Results: Full sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment (omitted: Non-Deterrence)										
Deterrence	0.185*		0.302**	0.285**	0.242**	0.303**	0.217**	0.288***	0.275**	0.210**
	(0.0869)		(0.0991)	(0.0833)	(0.0688)	(0.0950)	(0.0699)	(0.0760)	(0.0907)	(0.0659)
Delivery (omitted: Physical Letter)										
Digital Letter	0.119			-0.0768						0.155*
	(0.141)			(0.177)						(0.0627)
In Person	0.128			0.253*						0.235
	(0.154)			(0.0990)						(0.119)
Tax Type (omitted: Corporate Tax)										
Income Tax	0.972***				0.177					0.0770
	(0.100)				(0.112)					(0.160)
Property Tax	1.312***				-0.131					0.0344
	(0.172)				(0.0772)					(0.180)
VAT	-0.0232				-0.384***					-0.214**
	(0.0964)				(0.0574)					(0.0711)
Other	0.369***				0.263*					0.246**
	(0.0908)				(0.107)					(0.0813)
All	0.194				-0.167					0.107
	(0.189)				(0.218)					(0.137)
Response Horizon (omitted: Long Run)										
Short Run	0.346***					0.294**				0.263**
	(0.0887)					(0.0974)				(0.0863)
Taxpayer Type (omitted: Individual)										
Business	0.757***						-0.0413			0.0557
	(0.134)						(0.0682)			(0.174)
Individual and Business	0.904***						-0.0905			-0.106
	(0.130)						(0.133)			(0.176)
Late-payer sample (omitted: On Time)										
Late	0.127**						0.388***			0.285***
	(0.0463)						(0.0981)			(0.0663)
Method (omitted: OLS)										
NLPM	0.141							-0.0424		0.102
	(0.230)							(0.137)		(0.103)
LPM	0.0263							0.0561		0.112
	(0.127)							(0.131)		(0.107)
DiD	-0.0208							-0.157		-0.0152
	(0.0213)							(0.130)		(0.114)
2SLS	-0.0941							0.0851		-0.124
	(0.133)							(0.195)		(0.145)
WLS	-0.0470							0.365***		-0.0684
	(0.0664)							(0.0904)		(0.0698)
Other	-0.219*							-0.220*		-0.180*
	(0.0868)							(0.100)		(0.0699)
Baseline Comparison (omitted: No Info)										
Baseline Info	-0.953***							0.213		0.121
	(0.0543)							(0.175)		(0.0882)
Letter	-0.389***							-0.230*		-0.143
	(0.0620)							(0.111)		(0.0956)
Number of Observations										
ln(Number of Observations)	0.0314							-0.0173		0.0112
	(0.0323)							(0.0325)		(0.0131)
Publication Status (omitted: Unpublished)										
Published	0.287								-0.263*	-0.263***
	(0.181)								(0.128)	(0.0726)
Constant	-1.206***	0.535***	0.383***	0.384***	0.379***	0.119	0.283***	0.519	0.484***	-0.0977
	(0.237)	(0.0790)	(0.0865)	(0.0903)	(0.0638)	(0.0836)	(0.0757)	(0.276)	(0.0882)	(0.213)
Study fixed effects	Yes	No	No	No	No	No	No	No	No	No
Observations	933	949	949	949	949	935	944	947	949	933
R^2	0.491	0.000	0.092	0.120	0.228	0.122	0.228	0.172	0.153	0.385

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes : Regressions are estimated according to Equation 1. The dependent variable is the t-value of primary studies transformed to a dummy variable that equals 1 for estimates larger than 1.96 (i.e., that are positive and statistically significant at least at the 5% level) and 0 otherwise. All regressions are estimates with an OLS model. Column 1 includes study fixed effects. The independent variables of interest are a set of categorical variables entering the regression one by one. The ommited category is noted at each sub-heading and is emphasized with bold letters.

the magnitudes of these effects are generally so small that the experiments do not have the power to discriminate between the magnitudes of short- and long-run effects.

Second, we find that a key feature of the experimental design, its delivery method, in general matters for compliance. The t-value analysis of Table 4 suggests that interventions delivered by in-person visits to taxpayers are more likely to find a significant treatment effect than interventions delivered through letters. This effect is not significant in all models in a robust way, however Table A2 suggests that the magnitude of compliance measures at both extensive and intensive margins is substantially higher during personal visits by tax auditors as compared to treatments delivered by physical letters.

Third, regarding the characteristics of taxpayers, in Table 4 we find that nudges are more effective when addressing sub-sample of taxpayers who missed their deadline of paying taxes as well as to the those addressing VAT payers. The evidence on late payers also holds within studies. Regarding VAT payers, in line with the evidence in Table 4, a nudge delivered to a VAT payer is 19% less likely to yield a significant treatment difference from the control group compared to a nudge delivered to a corporate income taxpayer (Column 10). The fact that VAT is less sensitive to nudging may be due to the low baseline evasion levels of VAT because of its self-enforcing properties (Naritomi 2019, Pomeranz 2015, Waseem 2019). For both late and VAT payers the direction of the point estimates as plotted in the reduced sub-samples in Table A2 go in a similiar direction as in t-value analysis of Table 4 but are typically not distinguishable from zero.

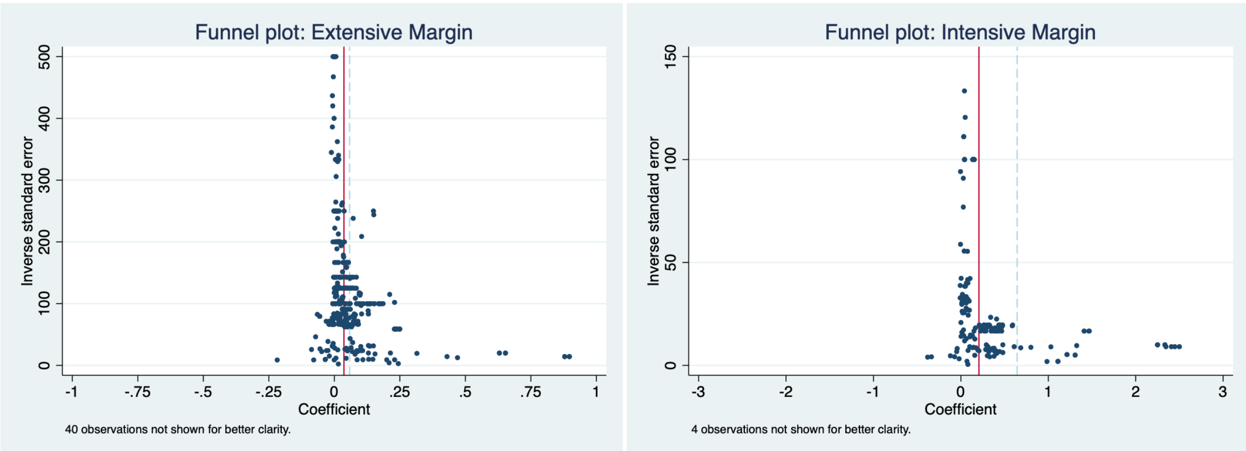
Finally, in Table A3 of the Appendix we interact these characteristics with the types of nudges to test whether the effects we find here hold generally for all nudge types or if they are mainly driven by deterrence nudges. The evidence of Table A3 in general does not support the hypothesis that the effects of these characteristics are driven by a particular type of a nudge.

5.4 Publication selection bias

One standard question often discussed in the meta-analytical literature is that the estimated treatment effects shown in the primary studies are systematically biased towards positive and significant effects. The underlying hypothesis is that researchers tend to present results that show: i) positive effects because it is generally believed that nudges should only have positive effects (file drawer bias), and ii) statistically significant effects because of the belief that non-significant effects are harder to publish (p-hacking).

Figure 4: Funnel plots

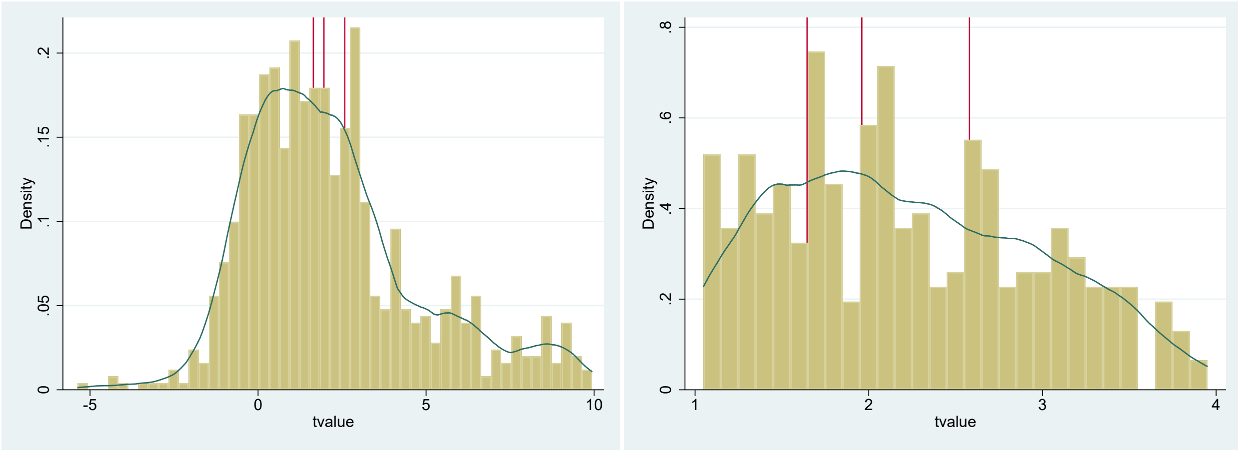
(a) Extensive margin **(b)** Intensive margin



Notes : The red undashed (blue dashed) vertical line shows precision weighted (unweighted) mean of the treatment effect. For visual clarity, sub-figure (a) drops 40 outlier observations that are larger than 500 on the y-axis, and sub-figure (b) drops 4 obervations that are larger than 10 on the x-axis.

Funnel plots are a common way of visually diagnosing meta-datasets for the file drawer bias. These plots check for asymmetries in the relation between treatment effect magnitudes and measures of their precision. The idea is that absent publication bias very imprecise estimates should be randomly distributed around zero rather than being skewed to the right. We present funnel plots for our extensive and intensive margin samples in Figures 4 (a) and (b), where the x-axis plots the size of the treatment effect and the y-axis plots the inverse standard error of the treatment effects as a measure of precision. The red undashed and blue dashed vertical lines also show the precision weighted and unweighted means of the treatment effect. In both samples we observe that the imprecisely estimated treatment effects, i.e.

Figure 5: Distribution of t-values in the full sample



$Notes$: Figures plot histograms of t-values of treatment effects which we obtained from the primary studies in our sample. For visual clarity, the left sub-figure drops outlier observations that lie outside the (-10, 10) range. The right sub-figure plots the same data as in the left sub-figure but zooming in to the range (1 , 4). Vertical lines denote critical values for two-sided significance tests at t-values of 1.645, 1.96 and 2.58. The kernel density line is estimated according to an Epanechnikov function.

those in the bottom of the funnel plot, tend to be skewed towards positive values. This visual evidence speaks for the presence of file drawer type bias in our sample.

One approach used to study p-hacking type of bias is to check for unusual patterns around critical values in the distribution of t-values. Such evidence is presented by Brodeur et al. (2016). The paper uses a large data comprising of tests published in top economics journals, and shows a disproportionately large share of tests that narrowly reject the null hypothesis. We follow this test and plot the distribution of t-values in Figure 5. We observe bunching in the number of observations of t-values situated just right to the three critical values which are denoted by vertical lines. We also observe corresponding missing masses on the left sides of the critical values. This evidence suggests that part of the studies in our sample select to report results that are statistically significant at conventional level, and ignore treatment effect estimates that narrowly miss to reject the null hypothesis.

A third and related idea is that, in addition to researchers selecting to report stronger results in working papers, the similar preferences of journals further amplifies this selection during the publication process (see, e.g., Andrews and Kasy 2019). This leads to the testable hypothesis that biased significant and positive treatment effects in working paper versions of studies will tend to become even more biased in published versions. We test this hypothesis in column 9 and 10 of Table 4 by comparing the average differences in significance levels between working papers and published papers. Table A2 of the Appendix extends the analysis to the sub-samples measuring responses at the extensive and intensive margins. In contrast to the view that the peer-review process exacerbate the publication selection bias, Tables 4 shows that published papers present results that are 30% less likely to be statistically significant than results shown in working papers. In terms of the magnitudes of treatment effects, Table A2 does not show evidence that the effects sizes are different between published papers and working papers either at the extensive or intensive margins.[31] The negative and significant point estimates on the publication status dummy suggests that the peer-review process tends to serve as a check against the incentives of researchers to report results selectively.

This evidence for selective reporting of results that we find is similiar to the findings of many other meta-analytical applications in economics. This suggests that empirical studies implementing RCTs, which are otherwise believed to have relatively sound methodologies, are not immune to biased reporting of results.[32]

6 Conclusions

Policy interventions that nudge taxpayers with the aim of increasing compliance have become an attractive tool among many governments due to their ease of implementation and low monetary costs. This easy adoption of the policy is demonstrated, for example, by Hjort et al. (2019) who inform Brazilian mayors about research on the positive tax compliance effects of reminder letters in an experimental setting and find that the treated municipalities are more likely to implement nudging interventions. However, little is known about the effectiveness

[31]We have additionally included a categorical variable in these models that measures the quality of the journal publishing the paper (distinguishing between top five, top field, second tier field and other journals). We did not find evidence for differences in either the significance or size of treatment effects between these journals of different quality.

[32]See Brodeur et al. (2018) for evidence on how publication selection bias differs by the identification method used.

of nudges beyond the evidence presented in individual experiments carried out in different contexts.

In this paper we summarize the knowledge accumulated so far from over 40 nudging interventions in a systematic way. We show that, unlike the general excitement over nudges in policy and academic circles, communications informing taxpayers about the morale aspects of paying taxes are not very effective in increasing compliance. Although, nudges that threaten taxpayers with audit probabilities and other elements of deterrence can be effective, the magnitudes of these effects are fairly small and are likely to be bound to the short-run.

Our evidence in general warns against the widespread and unconditional adoption of tax nudges in practice. However, this is not to say that nudges are useless. Our evidence on the particular design features of interventions that make them more effective (e.g., sending deterrence rather than only non-deterrence letters) combined with the identification of the sub-populations of taxpayers that are likely to be more sensitive to nudges (e.g., focusing on late-payers) provide guidance for potentially more effective policy interventions in the future. Note that the nudges we study are arguably the most common types of behavioral interventions, but governments can nudge in other ways too. For example, policies that publicly recognize the top taxpayers and shame the tax delinquents, as studied by Slemrod et al. (2019) and Dwenger and Treber (2018), or ones that use third-party information reports to pre-fill tax returns, as studied by Fochmann et al. (2018), Gillitzer and Skov (2018), Kotakorpi and Laamanen (2016), might as well be considered as nudges in a broader sense of the word.

This review also highlights a number of opportunities for researchers by directing attention towards gaps in the literature where the evidence has been weak so far. For example, only few papers test whether nudges work in the longer run, and when implemented repeatedly. Evidence on the question of whether the strength of deterrence (e.g., different audit probabilities or fine rates) and non-deterrence (e.g., different degrees of public goods) nudges matters is also lacking. Importantly, we do not have much knowledge on whether interventions interact with the context they operate in. This is not surprising given that randomized control trials tend to narrowly focus on local environments where the context is fixed. Cross-study comparisons such as the one adopted in this paper, on the other hand, are limited due to methodological concerns in comparing different experiments. Such an analysis in our paper would be additionally constrained due to the fact that interventions so far have mainly focused on Europe and the Americas leaving us with little cross-sectional variation to exploit. Future interventions, possibly ones that span across borders, could try to study i) whether non-deterrence nudges work more effectively in contexts of higher levels of trust, and ii) if deterrence nudges work better in uncorrupt environments where audits can be enforced more credibly compared to institutionally less mature environments.

References

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–276.

Allingham, M. and A. Sandmo (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics 1*(3-4), 323–338.

Alm, J. (2019). What motivates tax compliance? *Journal of Economic Surveys 33*(2), 353–388.

Alm, J., A. Malezieux, and M. McKee (2018). 40 years of tax evasion games: A meta-analysis. Working paper.

Almunia, M., I. Guceri, B. Lockwood, and K. Scharf (2019). More giving or more givers? the effects of tax incentives on charitable donations in the uk. *Journal of Public Economics, forthcoming*.

Anderson, R. G. and A. Kichkha (2017). Replication, meta-analysis, and research synthesis in economics. *American Economic Review 107*(5), 56–59.

Andreoni, J., B. Erard, and J. Feinstein (1998). Tax compliance. *Journal of economic literature 2*(36), 818–860.

Andrews, I. and M. Kasy (2019). Identification of and correction for publication bias. *American Economic Review 109*(8), 2766–2794.

Appelgren, L. (2008). The effect of audit strategy information on tax compliance-an empirical study. *eJournal of Tax Research 6*(1), 67–81.

Ariel, B. (2012). Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial. *Criminology 50*(1), 27–69.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 2*(76), 169–217.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon, and S. Galing (2017). Should governments invest more in nudging? *Psychological science 28*(8), 1041–1055.

Bérgolo, M. L., R. Ceni, G. Cruces, M. Giaccobasso, and R. Perez-Truglia (2017, July). Tax audits as scarecrows: Evidence from a large-scale field experiment. Working Paper Series 23631, National Bureau of Economic Research.

Bernheim, B. D. and D. Taubinsky (2018). Behavioral public economics. In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 1, pp. 381–516. Elsevier.

Besley, T., A. Jensen, and T. Persson (2019). Norms, enforcement, and tax evasion. Technical report, National Bureau of Economic Research Working Paper No. 25575.

Biddle, N., K. M. Fels, and M. Sinning (2018). Behavioral insights on business taxation: Evidence from two natural field experiments. *Journal of Behavioral and Experimental Finance 18*, 30–49.

Blackwell, C. (2007). A meta-analysis of tax compliance experiments. In Martinez-Vazquez and J. Alm (Eds.), *Tax Compliance and Evasion*.

Blumenthal, M., C. Christian, J. Slemrod, and M. G. Smith (2001). Do normative appeals affect tax compliance? evidence from a controlled experiment in minnesota. *National Tax Journal 54*(1), 125–138.

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. In *Handbook of labor economics*, Volume 3, pp. 1559–1695. Elsevier.

Boning, W. C., J. Guyton, R. H. Hodge, J. Slemrod, U. Troiano, et al. (2018). Heard it through the grapevine: direct and network effects of a tax enforcement field experiment. Technical report, National Bureau of Economic Research Working Paper No. 24305.

Bott, K. M., A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden (2017). You've got mail: A randomised field experiment on tax evasion. Discussion Paper 10/2017, Norwegian School of Economics, Department of Economics.

Boyer, P. C., N. Dwenger, and J. Rincke (2016). Do norms on contribution behavior affect intrinsic motivation? field-experimental evidence from germany. *Journal of Public Economics 144*, 140 – 153.

Brockmeyer, A., M. Hernandez, S. Kettle, and S. Smith (2019). Casting a wider tax net: Experimental evidence from costa rica. *American Economic Journal: Economic Policy*.

Brodeur, A., N. Cook, and A. Heyes (2018). Methods matter: P-hacking and causal inference in economics and finance. *IZA DP No. 11796*.

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics 8*(1), 1–32.

Card, D., J. Kluve, and A. Weber (2010). Active labour market policy evaluations: A meta-analysis. *The Economic Journal 120*(548), F452–F477.

Card, D., J. Kluve, and A. Weber (2017). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association 16*(3), 894–931.

Castro, L. and C. Scartascini (2015). Tax compliance and enforcement in the pampas evidence from a field experiment. *Journal of Economic Behavior & Organization 116*, 65 – 82.

Chirico, M., R. Inman, C. Loeffler, J. MacDonald, H. Sieg, J. A. Mortenson, H. R. Schramm, A. Whitten, D. Shoag, C. Tuttle, et al. (2019). Deterring property tax delinquency in philadelphia: An experimental evaluation of nudge strategies. *National Tax Journal 72*(3), 479–506.

Coleman, S. (1996). The minnesota income tax compliance experiment: State tax results.

Costa, D. L. and M. E. Kahn (2013). Energy conservation “nudges” and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*(3), 680–702.

De Neve, J.-E., C. Imbert, M. Luts, J. Spinnewijn, and T. Tsankova (2019). How to improve tax compliance? evidence from population-wide experiments in belgium. Technical report, CEPR Discussion Paper 13733.

Del Carpio, L. (2013). Are the neighbors cheating? evidence from a social norm experiment on property taxes in peru. Working paper, Princeton University.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review 109*(8), 2728–65.

Doerrenberg, P. and J. Schmitz (2017). Tax compliance and information provision. a field experiment with small firms. *Journal of Behavioral Economics for Policy 1*(1), 47–54.

Dwenger, N., H. Kleven, I. Rasul, and J. Rincke (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in germany. *American Economic Journal: Economic Policy 8*(3), 203–32.

Dwenger, N. and L. Treber (2018). Shaming for tax enforcement: Evidence from a new policy. Technical report, CEPR Discussion Papers No. 13194.

Eerola, E., T. Kosonen, T. Lyytikäinen, and J. Tuimala (2019). Tax compliance in the rental housing market: Evidence from a field experiment. Technical report, VATT Institute for Economic Research Working Papers No 122.

Farhi, E. and X. Gabaix (2019). Optimal taxation with behavioral agents. *The American Economic review, forthcoming*.

Feld, L. P. and J. H. Heckemeyer (2011). FDI and taxation: A meta-study. *Journal of economic surveys 25*(2), 233–272.

Fellner, G., R. Sausgruber, and C. Traxler (2013, 06). Testing enforcement strategies in the field: Threat, moral appeal and social information. *Journal of the European Economic Association 11*(3), 634–660.

Fochmann, M., N. Müller, and M. Overesch (2018). Less cheating? the effects of prefilled forms on compliance behavior. *Arqus Discussion Paper No. 227*.

Gechert, S. (2015). What fiscal policy is most effective? a meta-regression analysis. *Oxford Economic Papers 67*(3), 553–580.

Gillitzer, C. and M. Sinning (2018). Nudging businesses to pay their taxes: Does timing matter? Iza discussion paper 11599.

Gillitzer, C. and P. E. Skov (2018). The use of third-party information reporting for tax deductions: evidence and implications from charitable deductions in Denmark. *Oxford Economic Papers 170*(3), 892–916.

Hallsworth, M., J. A. List, R. D. Metcalfe, and I. Vlaev (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics 148*, 14 – 31.

Harju, J., T. Kosonen, and O. Ropponen (2018). Do honest hairdressers get a haircut. *Unpublished manuscript*.

Hasseldine, J., P. Hite, S. James, and M. Toumi (2007). Persuasive communications: Tax compliance enforcement strategies for sole proprietors. *Contemporary Accounting Research 24*(1), 171–194.

Heinemann, F., M.-D. Moessinger, and M. Yeter. (2018). Do fiscal rules constrain fiscal policy? a meta-regression-analysis. *European Journal of Political Economy 51*, 69–92.

Hernandez, M., J. Jamison, E. Korczyc, N. Mazar, and R. Sormani (2017). Applying behavioral insights to improve tax collection - experimental evidence from poland. Working paper, The World Bank.

Hiscox, M. Improved compliance with the deferred gst scheme. *Behavioural Economics Team of the Australian Government, Working Paper*.

Hjort, J., D. Moreira, G. Rao, and J. F. Santini (2019). How research affects policy: Experimental evidence from 2,150 brazilian municipalities. Technical report, National Bureau of Economic Research Working Paper No. 25941.

Iyer, G. S., P. M. Reckers, and D. L. Sanders (2010). Increasing tax compliance in washington state: A field experiment. *National Tax Journal 63*(1), 7.

John, P. and T. Blume (2018). How best to nudge taxpayers? the impact of message simplification and descriptive social norms on payment rates in a central london local authority. *Journal of Behavioral Public Administration 1*(1), 1–11.

Kettle, S., M. Hernandez, S. Ruda, and M. Sanders (2016). Behavioral interventions in tax compliance: Evidence from guatemala. Policy research working papers 7690, The World Bank.

Kettle, S., M. Hernandez, M. Sanders, O. Hauser, and S. Ruda (2017). Failure to captcha attention: Null results from an honesty priming experiment in guatemala. *Behavioral Sciences 7*(2), 1–21.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Kleven, H. J. and C. T. Kreiner (2006). The marginal cost of public funds: Hours of work versus labor force participation. *Journal of Public Economics 90*(10-11), 1955–1973.

Kotakorpi, K. and J.-P. Laamanen (2016). Prefilled income tax returns and tax compliance: Evidence from a natural experiment. *University of tampere Working Paper No. 104*.

Lichter, A., A. Peichl, and S. Siegloch (2015). The own-wage elasticity of labor demand: A meta-regression analysis. *European Economic Review 80*, 94–119.

Luttmer, E. F. and M. Singhal (2014). Tax morale. *Journal of Economic Perspectives 28*(4), 149–68.

Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys 32*(2), 273–301.

Mascagni, G., A. T. Mengistu, and F. B. Woldeyes (2018). Can icts increase tax? experimental evidence from ethiopia. ICTD Working Paper 82, The International Centre for Tax and Development.

Mascagni, G., C. Nell, and N. Monkam (2017). One size does not fit all: A field experiment on the drivers of tax compliance and delivery methods in rwanda. ICTD Working Paper 58, The International Centre for Tax and Development.

Meiselman, B. S. (2018). Ghostbusting in detroit: Evidence on nonfilers from a controlled field experiment. *Journal of Public Economics 158*, 180 – 193.

Naritomi, J. (2019). Consumers as tax auditors. *American Economic Review 109*(9), 3031–3072.

Neisser, C. (2017). The elasticity of taxable income: A meta-regression analysis. Discussion paper no. 17-032, ZEW.

Nelson, J. P. and P. E. Kennedy (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment. *Environmental and resource economics 42*(3), 345–377.

Ortega, D. and P. Sanguinetti (2013). Deterrence and reciprocity effects on tax compliance: Experimental evidence from venezuela. CAF Working Papers 08/2013, Development Bank Of Latinamerica.

Ortega, D. and C. Scartascini (2015). Don't blame the messenger: A field experiment on delivery methods for increasing tax compliance. IDB Working Paper Series 627, Inter-American Development Bank.

Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics 167*, 120–137.

Pomeranz, D. (2015). No taxation without information: Deterrence and self-enforcement in the value added tax. *American Economic Review 105*(8), 2539–2569.

Pomeranz, D. and J. Vila-Belda (2019). Taking state-capacity research to the field: Insights from collaborations with tax authorities. *Annual Review of Economics 11*, 755–781.

Scartascini, C. and E. Castro (2019). Imperfect attention in public policy: A field experiment during a tax amnesty in Argentina. Technical report, IDB Discussion Paper No 665.

Shimeles, A., D. Z. Gurara, and F. Woldeyes (2017). Taxman's dilemma: Coercion or persuasion? evidence from a randomized field experiment in ethiopia. *American Economic Review 107*(5), 420—424.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives 21*(1), 25–48.

Slemrod, J. (2019). Tax compliance and enforcement. *Journal of Economic Literature 57*(4), 904–954.

Slemrod, J., M. Blumenthal, and C. Christian (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., O. U. Rehman, and M. Waseem (2019). Pecuniary and non-pecuniary motivations for tax compliance: Evidence from Pakistan. Technical report, National Bureau of Economic Research Working Paper No. 25623.

Slemrod, J. and S. Yitzhaki (2002). Tax avoidance, evasion, and administration. In *Handbook of public economics, Vol. 3*, Volume 2, pp. 1423–1470. Elsevier.

Stanley, T., H. Doucouliagos, M. Giles, J. Heckemeyer, R. Johnston, P. Laroche, J. Nelson, M. Paldam, J. Poot, G. Pugh, and R. Rosenberger (2013). Meta-analysis of economics research reporting guidelines. *Journal of economic surveys 27*(2), 390–394.

Stanley, T. D. (2001). Wheat from chaff: Meta-analysis as quantitative literature review. *Journal of economic perspectives 15*(3), 131–150.

Stanley, T. D. and H. Doucouliagos (2012). *Meta-regression analysis in economics and business*. New York & London: Routledge.

Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy 37*(4), 583–588.

Thaler, R. H. and C. R. Sunstein (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven & London: Yale University Press.

Torgler, B. (2004). Moral suasion: An alternative tax policy strategy? evidence from a controlled field experiment in switzerland. *Economics of Governance 5*(3), 235–253.

Waseem, M. (2019). Information, asymmetric incentives, or withholding? understanding the self-enforcement of value-added tax. Technical report, Mimeo.

Wenzel, M. (2006). A letter from the tax office: Compliance effects of informational and interpersonal justice. *Social Justice Research 19*(3), 345–364.

Wisdom, J., J. S. Downs, and G. Loewenstein (2010). Promoting healthy choices: Information versus convenience. *American Economic Journal: Applied Economics 2*(2), 164–78.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of public economics* (3), 201–202.

Appendix

Table A1: Summary statistics

	Full sample		Extensive margin		Intensive margin	
Variable	Obs	Mean	Obs	Mean	Obs	Mean
Treatment effect size			456	0.059	172	0.648
Treatment effect significance dummy	949	0.535	456	0.634	172	0.698
Nudge type (deterrence=0, non-deterrence=1)	949	0.495	456	0.5	172	0.314
Delivery method	949	0.400	456	0.338	172	0.395
Tax type	949	1.525	456	1.585	172	1.895
response horizon	935	0.889	452	0.887	172	0.849
Taxpayer type	949	0.732	456	0.649	172	0.895
Late payer sample	944	0.449	456	0.581	172	0.616
Method	949	2.299	456	2.355	172	2.238
Baseline comparison	947	1.320	456	1.355	172	1.494
Log number of observation	949	10.126	456	10.156	172	9.979
Publication status	949	0.331	456	0.377	172	0.221

Table A2: Heterogenous Results: Robustness at Extensive and Intensive Samples

	Extensive				Intensive			
	significance		coefficient		significance		coefficient	
Treatment (omitted: Non-Deterrence)								
Deterrence	0.266*	0.238**	0.0247***	0.00616	0.289	0.313	0.137*	0.162*
	(0.100)	(0.0774)	(0.00586)	(0.0106)	(0.179)	(0.164)	(0.0497)	(0.0666)
Delivery (omitted: Physical Letter)								
Digital Letter	0.0558	0.158	0.0973***	0.0377	0.136	0.376	0.109	-0.118
	(0.0941)	(0.0982)	(0.0237)	(0.0242)	(0.160)	(0.187)	(0.541)	(0.165)
In Person	0.115	0.154	0.198***	0.149**	0.205	0.457*	2.640**	2.003**
	(0.139)	(0.115)	(0.0313)	(0.0523)	(0.240)	(0.202)	(0.812)	(0.581)
Tax Type (omitted: Corporate Tax)								
Income Tax	0.327*	-0.307	-0.0928***	0.0574	3.152**	0.296	0.843	0.543
	(0.118)	(0.221)	(0.0165)	(0.0451)	(0.886)	(0.197)	(0.706)	(0.280)
Property Tax	0.321	-0.505*	-0.103***	0.0775	5.595**	0.452	0.604	-0.00635
	(0.163)	(0.212)	(0.0150)	(0.0381)	(1.844)	(0.298)	(1.614)	(0.412)
VAT	-0.110	-0.450***	-0.0958	-0.000979	2.159*	-0.482**	0.750	0.867*
	(0.236)	(0.0897)	(0.0587)	(0.0201)	(0.893)	(0.137)	(0.711)	(0.347)
Other	0.533**	-0.0486	0.183***	0.0346	2.493*	-0.0588	0.693	1.187***
	(0.183)	(0.0716)	(0.0296)	(0.0392)	(0.893)	(0.105)	(0.711)	(0.234)
All	-0.110	-0.466*	-0.0982	-0.0215	2.493*	-0.130	0.688	0.880**
	(0.236)	(0.176)	(0.0587)	(0.0376)	(0.893)	(0.117)	(0.711)	(0.288)
Response Horizon (omitted: Long Run)								
Short Run	0.380**	0.418**	0.0381	0.0338	0.303	0.406*	-0.601	-0.685
	(0.115)	(0.116)	(0.0261)	(0.0234)	(0.154)	(0.153)	(0.443)	(0.566)
Taxpayer Type (omitted: Individual)								
Business	0.250*	-0.272	-0.110***	0.0749	2.255*	0.293	0.628	-0.0105
	(0.0989)	(0.223)	(0.0169)	(0.0509)	(0.800)	(0.267)	(0.639)	(0.416)
Individual and Business	0.568***	-0.287	-0.112***	0.0345	2.945**	0.267	-0.178	-1.010
	(0.132)	(0.241)	(0.0257)	(0.0604)	(0.966)	(0.232)	(1.017)	(1.023)
Late-payer sample (omitted: On Time)								
Late	0.105	0.145	0.00202	0.0462	0.242	-0.0946	0.916	0.778**
	(0.0815)	(0.0941)	(0.00528)	(0.0229)	(0.266)	(0.290)	(0.587)	(0.217)
Method (omitted: OLS)								
NLPM	-0.130	0.0580	0.00179	-0.0298	1.691***	0.223	-0.500	-0.459
	(0.0799)	(0.103)	(0.00461)	(0.0149)	(0.269)	(0.346)	(0.450)	(0.700)
LPM	-0.129	0.0148	0.00159	0.0240	1.629***	0.545**	0.0844	-0.0638
	(0.0798)	(0.116)	(0.00449)	(0.0195)	(0.252)	(0.141)	(0.151)	(0.242)
DiD	-0.463**	-0.210	0.100**	0.0315	0.106	-0.109	0.0129	0.0924
	(0.126)	(0.390)	(0.0333)	(0.0385)	(0.0956)	(0.219)	(0.0737)	(0.196)
2SLS	-0.0903	-0.107	0.105	0.0992	0.0421	0.0562	2.376	2.385
	(0.0696)	(0.0699)	(0.0730)	(0.0689)	(0.0349)	(0.0356)	(1.504)	(1.422)
WLS	-0.0452	-0.0796	0.0530	0.0461	0.0210	0.0393	1.195	1.111
	(0.0348)	(0.0652)	(0.0365)	(0.0352)	(0.0175)	(0.0317)	(0.752)	(0.682)
Baseline Comparison (omitted: No Info)								
Baseline Info	-0.974***	0.0520	-0.0120**	0.0168	-0.737**	0.330	0.0724	-0.366
	(0.0159)	(0.0900)	(0.00424)	(0.0166)	(0.203)	(0.219)	(0.141)	(0.333)
Letter	-0.207	-0.246	-0.0110	0.00971	-0.396	-0.341	0.0800	-0.147
	(0.112)	(0.217)	(0.0105)	(0.0292)	(0.288)	(0.229)	(0.184)	(0.253)
Publication Status (omitted: Unpublished)								
Published	-0.714***	-0.268*	0.000266	0.00937	1.581	-0.197	0.697	-0.253
	(0.0869)	(0.0968)	(0.00572)	(0.0131)	(1.093)	(0.118)	(0.866)	(0.233)
Number of Observations								
ln(Observations)	-0.0646	-0.0361	-0.00748	-0.0152***	-0.160	0.0502	-0.0914	0.0419
	(0.0390)	(0.0272)	(0.0104)	(0.00400)	(0.145)	(0.0245)	(0.111)	(0.0500)
Constant	0.776*	0.857**	0.168	0.0551	-3.841***	-0.695**	-0.137	-0.467
	(0.321)	(0.298)	(0.109)	(0.0517)	(0.704)	(0.220)	(0.619)	(0.814)
Paper FE	Yes	No	Yes	No	Yes	No	Yes	No
Observations	452	452	452	452	172	172	172	172
R^2	0.541	0.453	0.465	0.396	0.642	0.499	0.419	0.400

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A3: Interactions

	(1) Extensive	(2) Extensive	(3) Intensive	(4) Intensive	(5) Full	(6) Full
Treatment (omitted: Non-Deterrence)						
Deterrence	0.0242	0.00755	-2.292**	-3.037***	0.161	-0.525
	(0.0565)	(0.0276)	(0.652)	(0.311)	(0.342)	(0.439)
Delivery (omitted: Physical Letter)						
Digital Letter	-0.0252	0.127***	0.141	0.571***	0.120	0.460*
	(0.0372)	(0.0165)	(0.253)	(1.06e-10)	(0.263)	(0.207)
In Person	-0.0469	0.0967***	0.848***	0.793	0.425	0.565**
	(0.0538)	(0.0167)	(0.0723)	(.)	(0.262)	(0.182)
Tax Type (omitted: Corporate Tax)						
Income Tax	0.145**	0.142**	0.215	2.789***	0.134	0.383
	(0.0486)	(0.0491)	(0.160)	(0.203)	(0.224)	(0.274)
Property Tax	0.106*	0.226*	0.779**	4.850***	-0.0105	1.306***
	(0.0384)	(0.0870)	(0.261)	(0.100)	(0.224)	(0.293)
VAT	0.00328	-0.0771	0.146	0.205	-0.297***	0.0271
	(0.0251)	(0.0476)	(0.0836)	(0.114)	(0.0467)	(0.0374)
Other	0.128**	0.157***	0.0930	0.137	0.0103	-0.0376
	(0.0384)	(0.0174)	(0.0677)	(0.115)	(0.165)	(0.0963)
All	-0.00915	-0.0813	0.110	0.154	-0.156	0.00474
	(0.0376)	(0.0479)	(0.0677)	(0.115)	(0.121)	(0.0591)
Response Horizon (omitted: Long Run)						
Short Run	0.0298*	0.0285**	0.0509	0.104	0.240*	0.356***
	(0.0122)	(0.00947)	(0.111)	(0.0928)	(0.0900)	(0.0884)
Taxpayer Type (omitted: Individual)						
Business	0.206***	0.137**	0.235	2.370***	0.0520	0.155
	(0.0273)	(0.0455)	(0.170)	(0.203)	(0.224)	(0.272)
Individual and Business	0.128**	0.135*	0.340	2.713***	-0.262	0.192
	(0.0382)	(0.0526)	(0.177)	(0.228)	(0.225)	(0.285)
Late-payer sample (omitted: On Time)						
Late	0.0216	0.0150	0.375***	-2.415***	0.250*	0.122
	(0.0225)	(0.0124)	(0.0367)	(0.0210)	(0.102)	(0.0685)
Baseline Comparison (omitted: No Info)						
Baseline Info	-0.0538*	-0.0151***	0.0681	-0.0518	0.181	-0.828***
	(0.0211)	(6.59e-10)	(0.106)	(0.0601)	(0.116)	(0.154)
Letter	-0.0593	0.0222	0.486	0.254*	-0.0867	0.139
	(0.0318)	(0.0179)	(0.300)	(0.106)	(0.161)	(0.219)
Middle Income	0.00825	0.0892	0.383***	0.458***	-0.0906	0.447*
	(0.0392)	(0.0533)	(0.0839)	(0.0875)	(0.290)	(0.213)
High Income	0.0295	0.0683	0.618***	2.793***	-0.119	0.424
	(0.0378)	(0.0527)	(0.0277)	(3.51e-10)	(0.262)	(0.352)
Interactions						
Deterrence × Digital Letter	0.0727	-0.0381*	0	0	-0.117	-0.390
	(0.0441)	(0.0176)	(.)	(.)	(0.255)	(0.232)
Deterrence × In Person	0.243***	0.127***	1.967***	2.540***	-0.467	-0.540**
	(0.0580)	(0.0195)	(0.378)	(1.01e-09)	(0.266)	(0.190)
Deterrence × Income Tax	-0.0862	-0.0365	1.147*	0.216	-0.0939	0.764*
	(0.0481)	(0.0185)	(0.418)	(0.203)	(0.308)	(0.298)
Deterrence × Property Tax	-0.0427	-0.0443	0	0	0.0954	0.886**
	(0.0391)	(0.0345)	(.)	(.)	(0.283)	(0.304)
Deterrence × VAT	-0.0189	-0.0110	1.532**	2.703***	0.145	0.120
	(0.0334)	(0.0184)	(0.498)	(0.114)	(0.0980)	(0.0975)
Deterrence × Other	-0.0874*	-0.0902*	0.847**	2.719***	0.332	0.858***
	(0.0406)	(0.0407)	(0.247)	(0.117)	(0.192)	(0.179)
Deterrence × All	0.0413	-0.00540	1.417**	2.687***	0.572*	0.643*
	(0.0489)	(0.0241)	(0.398)	(0.117)	(0.231)	(0.246)
Deterrence × Short Run	0.0188	0.0387	-0.166	0.0889	-0.0819	-0.0226
	(0.0297)	(0.0271)	(0.255)	(0.117)	(0.135)	(0.143)
Deterrence × Business	-0.135***	-0.0384	1.090*	0.540*	-0.104	0.758*
	(0.0347)	(0.0211)	(0.384)	(0.203)	(0.318)	(0.290)
Deterrence × Individual and Business	-0.0415	0	0.953*	0.273	0.202	0.968**
	(0.0416)	(.)	(0.369)	(0.238)	(0.297)	(0.309)
Deterrence × Late	-0.00168	-0.0228	1.631**	2.980***	0.117	0.000129
	(0.0265)	(0.0249)	(0.422)	(2.29e-09)	(0.135)	(0.183)
Deterrence × Baseline Info	0.0480	-0.001000	0.238	0.165	-0.130	-0.130
	(0.0254)	(0.00894)	(0.280)	(0.114)	(0.137)	(0.164)
Deterrence × Letter	0.0530	-0.0234	0	0	-0.330*	-0.616**
	(0.0466)	(0.0163)	(.)	(.)	(0.161)	(0.185)
Deterrence × Middle Income	0.0373	0.0352*	-0.405	-0.308*	0.241	0.0390
	(0.0393)	(0.0157)	(0.280)	(0.114)	(0.284)	(0.302)
Deterrence × High Income	0.0291	0.0519**	1.098*	0	0.245	0.130
	(0.0405)	(0.0162)	(0.480)	(.)	(0.273)	(0.337)
Constant	-0.141***	-0.213*	-0.779**	-2.912***	0.0514	-0.942**
	(0.0357)	(0.0897)	(0.220)	(0.282)	(0.266)	(0.335)
Paper FE	No	Yes	No	Yes	No	Yes
Observations	452	452	172	172	924	924
R^2	0.401	0.458	0.360	0.368	0.385	0.515

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Download ZEW Discussion Papers from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/

or see:

https://www.ssrn.com/link/ZEW-Ctr-Euro-Econ-Research.html
https://ideas.repec.org/s/zbw/zewdip.html

//
IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de