ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 067

The Elasticity of Taxable Income: A Meta-Regression Analysis

Carina Neisser

February 2021

www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866 is gratefully acknowledged.

The Elasticity of Taxable Income: A Meta-Regression Analysis *

Carina Neisser

February 11, 2021

The elasticity of taxable income (ETI) is a key parameter in tax policy analysis. To examine the large variation found in the literature of taxable and broad income elasticities, I conduct a comprehensive meta-regression analysis using information from 61 studies containing 1,720 estimates. My findings reveal that estimated elasticities are not immutable parameters. They are correlated with contextual factors and the choice of the empirical specification influences the estimated elasticities. Finally, selective reporting bias is prevalent, and the direction of bias depends on whether deductions are included in the tax base.

JEL Classification: C81, H24, H26
Keyword: elasticity of taxable income; income tax; behavioural response; meta-regression analysis

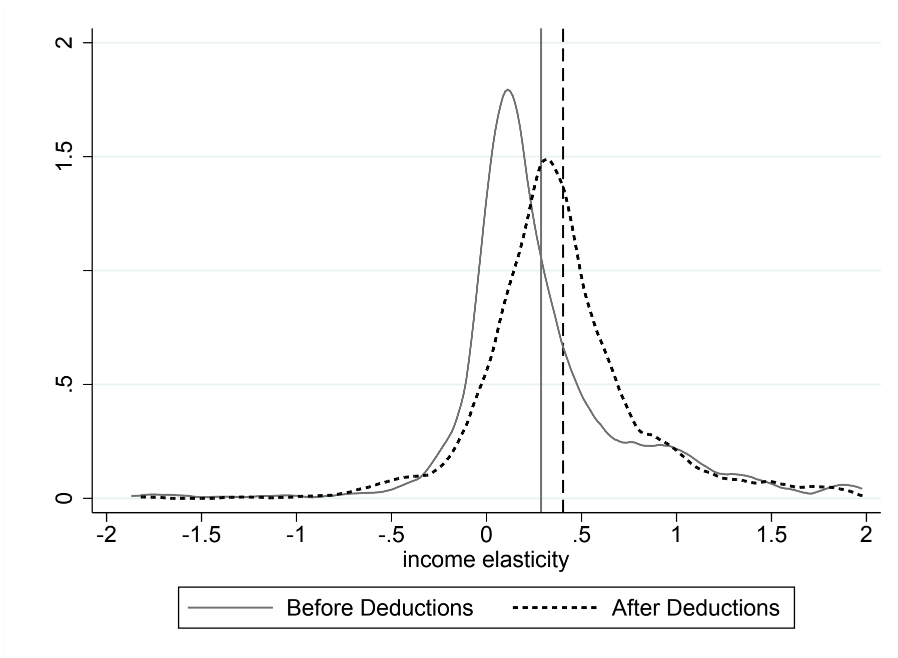
***Neisser**: University of Cologne and IZA Bonn (neisser@wiso.uni-koeln.de). Most of the work was done at ZEW - Leibniz Centre for European Economic Research Mannheim and I gratefully acknowledge financial support by Leibniz ScienceCampus MannheimTaxation. The project was also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866. I would like to thank David Agrawal, Philipp Dörrenberg, Paul Hufe, David Jaeger, Andreas Lichter, Max Löffler, Claus Kreiner, Andreas Peichl, Jörg Paetzold, Emmanuel Saez, Sebastian Siegloch, Holger Stichnoth, Christian Traxler, Johannes Voget, Mazhar Waseem, Mustafa Yeter and participants at the LAGV 2016, IIPF 2016, PF Seminar (U Mannheim), ZEW Public Finance 2017 and IEB Taxation Workshop 2017 for their helpful comments. Philipp Kollenda provided excellent research assistance.

The design of tax and transfer policies requires to quantify the magnitude of behavioural responses to tax rate changes in order to determine optimal policy. Larger responses to taxation, for example, will lead to smaller revenue-maximizing tax rates for top income earners, conditional on the shape of the income distribution (Saez, 2001; Saez et al., 2012). The elasticity of taxable income (ETI) summarizes different types of behavioural responses to income taxation such as real responses (e.g. labour supply adjustments), tax avoidance (e.g. (legally) claiming deductions or income shifting between tax bases) and illegal tax evasion behaviour. It serves not only as a behavioural parameter in optimal taxation models (Mirrlees, 1971; Diamond, 1998; Saez, 2001; Piketty and Saez, 2013) but also as sufficient statistic for dead-weight loss calculation (Feldstein, 1999 or Chetty, 2009). Since Feldstein (1995), a large body of empirical work estimating taxable income responses has emerged. Despite the importance, there is little consensus on the magnitude of these elasticities to be used in economic policy analysis.

In this paper, I provide a comprehensive quantitative survey by applying meta-regression techniques. 'Elasticity of taxable income' is used as an umbrella term for all types of elasticities (e.g. adjusted gross and taxable income). In total, I collect 1,720 estimates extracted from 61 studies. I only consider Difference-in-Differences (DID) and Instrumental Variable (IV) approaches and do not cover bunching (e.g. Saez, 2010) or time series evidence (e.g. Mertens and Montiel Olea, 2018) as these estimates are conceptually different and therefore not comparable to each other.

To account for the central role of deductions and to disentangle real and reporting responses by individuals, I explicitly differentiate between behavioural responses that are based on income concepts with or without tax deductions and allocate all reported elasticities to two subsamples: before (BD) and after deduction (AD) elasticities. In addition to real responses (e.g. changes in labour supply), many tax systems offer a wide range of deductions to legally avoid taxes. Figure 1 plots the distribution of elasticities of both subsamples. The vast majority of estimates (90%) lies within an interval of -1 and 1, with a strong propensity to report estimates between 0 and 1 and both distributions reveal an excess mass between 0.7 and 1. The broader range of responses is reflected by larger AD elasticities. Within my sample AD elasticities exhibit a mean of 0.403, while BD elasticities have a mean of 0.287.

Figure 1: Distribution of elasticities



Note: The distribution of before deduction (BD) elasticities are displayed as a solid line and the corresponding vertical line highlights the mean of 0.287 (N=940). The distribution of after deduction (AD) elasticities are displayed with a dashed line and the corresponding mean of 0.403 is highlighted with the vertical dashed line (N=780). Both figures are based on an an Epanechnikov kernel with a bandwidth of 0.072.

Researchers that estimate the ETI face various empirical challenges. Most importantly, income and marginal taxes are jointly determined and potential solutions like Instrumental Variables (IV) approaches have been developed. In addition, different income growth rates across the population or reversion to the mean require solutions because it is hard to disentangle income growth driven by tax and non-tax effects. Most notably functions of past income are included in the regressions. While the choice of the specific regression specification depends on the underlying model, there is some discretion in the way that specific methods and controls are implemented, which can partially affect results.

I identify and assess different explanations for the pattern of estimates found in the empirical literature. More precisely, different categories for each study (e.g. empirical strategy or country) are recorded and differences between elasticity estimates are quantitatively examined. Importantly, my meta-analysis provides a replicable statistical framework for summarizing and assessing the full range of empirical evidence.[1] Although the ETI literature has been reviewed by Saez et al. (2012), I am not aware of any meta-regression analysis of taxable income elasticity estimates.

My results show that elasticities that account for deductions are not only larger by definition, but they are also more sensitive to the estimation technique. A calculation of stylized elasticity estimates documents a wide range of possible estimates. When accounting for the implemented estimation specification in primary studies, my regression results show that average BD elasticities lie in the range of 0.053 to 0.120, while average AD elasticities vary from 0.074 to 0.827. Richer income control variables always lower estimated elasticities and the effect is more pronounced in the AD subsample. It remains unclear which income control is an appropriate choice to disentangle non-tax from tax-related income responses.

I link estimated elasticities to inequality measures as well as tax system- and economy related characteristics. More precisely, I add country and year specific characteristics to my collected data to to provide suggestive evidence that elasticities are related to contextual factors. Slemrod and Kopczuk (2002) and Kopczuk (2005) emphasise the fact that the ETI is considerably larger in tax systems with more deduction possibilities and can therefore be controlled by policy makers. Much of the evidence is based on self-employed and/or high-income taxpayers, given their larger range of opportunities to adjust their (taxable or gross) income (e.g. Kreiner et al., 2016; Le Maire and Schjerning, 2013 or Harju and Matikka, 2016). Alvaredo et al. (2013) highlight the role of tax policy and its effects on income inequality. In addition, Kleven et al. (2011) and Kleven et al. (2016) stress that third party information reporting (e.g. the exchange of information of employers or banks and tax authorities) influences the magnitude of behavioural responses.

My analysis provides evidence that estimated elasticities are not immutable parameters

[1] See Christensen and Miguel (2018) for a review of research transparency and reproducibility in economics. Card and Krueger (1995) and Card et al. (2010, 2018) are three examples that look into the field of labour economics. Havránek (2015) examines the literature on intertemporal substitution elasticities and Lichter et al. (2015) study labour demand elasticities.

with respect to the empirical strategy but they are also linked to past as well as current (tax-)policy and that the underlying context matters when interpreting these elasticities. There is a positive correlation between inequality measures and estimated elasticities. In particular, AD elasticities are highly correlated with top income shares. A widening of the income distribution might be the result of past tax cuts for high-income earners. Such developments are insufficiently considered in the initial estimation of elasticities of taxable income, leading to an upward bias in resulting AD elasticities. Moreover, the level of third party information reporting within an economy is unrelated to elasticities that account for the deduction component, while it is negatively related to the magnitude of elasticities that do not consider deductions. Typically, deductions are not subject to third party information reporting. The degree of information exchange between tax authorities and firms or other institutions can be influenced by policy makers and thereby also influence the magnitude of estimated elasticities .

I focus on two types of selective reporting bias. The first is the so-called 'file drawer problem' (Rosenthal, 1979). It describes the fact that many studies or results have never been published because they do not reveal the expected sign, magnitude and/or significance. The second type of selection reporting bias arises, if researchers use well-known results as a reference point and hence are inclined to report only results that are in line with these findings. With respect to the ETI, researchers generally put more trust into estimates ranging from 0 to 1. With their seminal contribution, Gruber and Saez (2002) have further shaped this belief by providing a value of 0.4 as their main result.

Graphical evidence as well as regression results confirm the prevalence of selective reporting bias in the literature of taxable income elasticities. In general, there is a tendency to report significant results more often. The existence of 'p-hacking' is more pronounced among AD elasticities and among published articles compared to working papers. Since the publication of Chetty (2009), BD (e.g. gross income) elasticities have begun to receive more attention. This increased interest is reflected by a larger amount of 'p-hacking' within the BD-subsample for estimates published after 2009. In addition, I observe excess (distributional) mass around 0 to 0.4 and below 1. These anomalies in the distribution of estimates suggests that results are more likely to get reported because they are in line with theory and existing evidence. In general, there is an upward reporting bias for BD elasticities. For AD elasticities, the reporting bias goes in both directions, while the downward bias appears to be more dominant.

The remainder of this paper is structured as follows. In Section 1, I explain the meta regression model and I describe the data collection process. In Section 2, I outline a basic framework to discuss empirical challenges in the literature on taxable income elasticities (2.1) and provide explanations of defined categories of heterogeneity (2.2) along with descriptive statistics (2.3). In Section 3, I provide and discuss the baseline results and correlations between contextual factors and elasticities. In Section 4, I highlight the prevalence of selective reporting bias. Section 5 concludes.

1 Meta-regression Framework and Data Collection

I follow standard meta-regression analysis techniques (e.g. Card et al., 2010, 2018). The meta regression model is given by

$$\zeta_{is} = \zeta_0 + \beta X_i + \delta Z_s + \epsilon_{is}, \tag{1}$$

where ζ_{is} represents the i-th estimate estimate collected from study s. ζ_0 denotes the intercept, X_i and Z_s represent study and estimate-specific variables respectively, and ϵ_{is} is the sampling error. Since the variances of collected estimates are heteroscedastic, it is preferable to estimate the model using Weighted Least Squares (WLS) rather than through an OLS estimation. I use the inverse of the error term variance of an individual estimate $\mathrm{V}(\hat{\zeta_{is}}) = \sigma_{is}^2$ as analytic weights. Hence, I give observations with smaller variances a larger weight and greater influence on the estimates since precision can be seen as an indicator of quality.[2] Standard errors are clustered at the study level to control for study dependence in

[2]To test the robustness of the results with respect to the underlying weights, I conduct various regressions (see (online) Appendix F): (1) a simple OLS, (2) Random effects meta-regression technique, (3) a WLS

the estimates.

Data Collection. A comprehensive review and examination of the ETI literature provided the data for the meta-analysis.[3] As a first step, I searched Google Scholar and IDEAS RePEc using the following search terms: 'elasticity of taxable income', 'eti', 'taxable income', 'new tax responsiveness' and 'tax elasticity.' In addition, I relied on a survey by Saez et al. (2012) to identify relevant studies published prior to 2011 and I cross-checked these with the reference list of all previously identified papers. I checked only English- and German-speaking articles. The main search process lasted from 2015 to 2019 and I identified 203 potential studies.

In the second step, I applied certain exclusion criteria to determine the final sample of studies. I only coded studies that provide their own empirical estimates and rely on commonly used income concepts as described below. Based on this sample, I found 37 studies that were published. Additional working papers increased the number of articles to 61.[4] In the third step, I collected every estimate derived from a different specification (so-called multiple sampling) so that they are different with regard to the defined categories of heterogeneity (e.g. income concept or sample restrictions). I collected all point estimates, corresponding standard errors, number of observations and type of control for heteroscedasticity and autocorrelation. Additional information on journal, year of publication, country and time period is coded. In the fourth step, I restricted the final dataset to estimates that provide a standard error or t-statistic. My sample consists of 1,720 observations. Finally, I collected all necessary study characteristics, which I will explain in the next section. Additional information on contextual factors such as tax system and economic characteristics as

with weights that are based on the inverse of the share of observations per study in relation to the full sample and (4) a WLS with weights that account for the sample size of each study. Last, to check whether clustering in the meta-analysis produces misleading inferences, I apply a wild-cluster bootstrap procedure for improved inference with only a few clusters.

[3]The meta analysis follows reporting guidelines proposed by Stanley et al. (2013). A list of people who have coded and checked the data, a list of identified but not-included studies and estimates or a list of all included estimates plus sources is provided upon request.

[4]In the (online) Appendix A, I provide an overview of studies included in the sample. On the one hand, adding unpublished papers to the meta-sample might lower the quality of included estimates but, on the other hand, most working papers are more recent and use better datasets and improved estimation techniques. It should be noted that this meta study is only as good as the studies on which it is based and there might be variation among the studies that cannot be reflected by the coded variables.

well as inequality measures are collected and merged with the dataset (see Table 1 for an overview).

2 Elasticity of Taxable Income

In this Section, I briefly explain the concept of taxable income elasticities. I outline the most standard regression specification and I state empirical challenges. For a detailed discussion, please refer to an excellent survey by Saez et al. (2012). I present various reasons why elasticity estimates differ and describe the coded characteristics along with a more in depth explanation in Section 2.2. In Section 2.3 I provide some descriptive statistics.

2.1 Empirical Challenges

The (taxable) income literature uses an extension of the traditional labour supply model. Individuals maximize a utility function $u(c, z)$, where z is income and c consumption. An elasticity of the income tax base measures the responsiveness of income to changes in the net-of-tax rate (NTR) - defined as one minus the marginal tax rate. This is the percentage change in income in response to a one percent increase in the NTR. An increase in the marginal tax rate reduces the NTR, which in turn reduces taxable income. Hence, the expected elasticity should be positive.[5]

Collected elasticity estimates are summarized such that they belong either to the before or after deductions subsample. Since an elasticity is a function of the definition of the tax base, the applied income concept determines the range of responses. These responses can take many forms, including changes in labour supply (participation and working hours), tax avoidance (changing the timing of income/transactions, changes in the extent of spending on tax deductible activities, e.g. donations, or even claiming questionable deductions) and tax evasion (understating income, claiming unjustified deductions). The distinction between whether or not an income concept considers deductions is crucial. Real responses can be

[5] Information about estimated income effects is rarely available (e.g. Gruber and Saez, 2002 or Bakos et al., 2010) so, I ignore them and assume that compensated and uncompensated elasticities are equal.

captured with a before-deduction elasticity while an after-deduction elasticity captures a broader range of responses, including avoidance behaviour. Tax evasion affects both types of elasticities. Ideally, we would like to observe a comparable and uniformly defined income across all studies. This is impossible even for conceptually equal income concepts like taxable income. The exact definition varies from country to country and, even within a country, over time. Researchers mainly use taxable, adjusted gross, or total income to capture behavioural responses towards taxation. Total income (= gross or broad income) is the sum of all income. Subtracting specific deductions (e.g. retirement plan contributions), yields adjusted gross income. Taxable income is calculated as adjusted gross income minus personal exemptions and itemized deductions.[6]

The most standard regression specification is derived as:

$$log\left(\frac{z_{it}}{z_{it-k}}\right) = \zeta log\left(\frac{1-\tau_{it}}{1-\tau_{it-k}}\right) + \delta f(z_{it-k}) + \theta X_{it-k} + \mu_t + \epsilon_{it}, \qquad (2)$$

where i refers to the respective taxpayer and t is the underlying year. ζ is the parameter of interest, k is the chosen difference length and $t-k$ denotes the base-year. X_{it-k} is a vector of control variables. Time dummies μ_t control for any omitted variables in differences that are the same on average for all individuals. $f(z_{it-k})$ denotes the income control in order to capture non-tax related income trends. In equation (2) ζ represents the elasticity of the income tax base that measures the responsiveness of income to changes in the net-of-tax rate $(1-\tau)$.

Several conditions must hold in order to estimate behavioural responses correctly. First, only the marginal tax rate τ should change, while changes in the tax base z are kept constant. In reality, however, the underlying tax base often changes simultaneously with the tax rate itself. To rule out any tax legislation-induced tax base effects, the broadest definition and, therefore, an 'artificial' tax base across years is used. For the US, researchers mostly rely on the TAXSIM calculator developed at the NBER. In other cases, the constant tax base along

[6]In the (online) Appendix B, Table 8 provides summary statistics by reported *income concept*. As a sensitivity check, I run the estimations on a subsample of the dataset and look only at taxable income elasticities (see Table 17). These results remain unchanged compared to estimation results that consider all AD estimates.

with a tax simulation model is often constructed by the researcher himself. Building a tax simulation model requires a broad understanding of the underlying tax law as well as tax base changes across the years under study.[7]

Second, in a progressive tax system, the marginal tax rate τ and income z are jointly determined, and tax rates increase automatically if an individual faces a (non-tax related) positive income shock and potential income responses are (wrongly) captured by the ETI. Following Auten and Carroll (1999) and Gruber and Saez (2002), most studies use an instrument that is based on mechanical changes in tax rates that are induced by tax reforms. The idea is that this change in net-of-tax rates is free of any behavioural responses, representing only mechanical changes that can be used as an instrument for the NTR. To construct mechanical tax rate changes, one uses income from base year $t - k$ and assumes that it remains the same in year t. Applying tax rules for year t yields a mechanical (sometimes called predicted or synthetic) tax rate. More developed instruments try to account for the growing concern that this instrument is not sufficiently exogenous. For instance, Weber (2014) argues that mechanical tax rate changes mentioned above should be lagged in order to fulfill the exclusion restriction. Her approach makes it possible to deal with serially correlated transitory income shocks.

Third, different income growth rates across the population (e.g. larger income growth for high-income earners) and reversion to the mean further aggravate a 'clean' estimation. For example, when income changes are driven by temporary income shocks or different parts of the income distribution grow at a higher rate, it is hard to disentangle income growth driven by tax and non-tax effects. In the case of tax cuts for upper-income groups, secular changes in income (e.g. larger income growth at the top), lead to an upward bias and mean reversion might go in both directions depending on the type of income shock. These shocks influence the shape of an income distribution and they need to be incorporated in an empirical framework. While administrative tax datasets offer precise information about a taxpayer's

[7] In many studies, details of the tax simulation model are missing. For example, although capital gains are part of taxable income, only a few studies explicitly mention that they subtract capital gains. In addition, most researchers remain salient on whether or not they apply a constant tax base approach. Both things influence the definition of taxable income and therefore the results.

income and deductions, socio-demographic information and therefore the amount of other control variables is limited. To capture non-tax related income growth, researchers use income controls $f(z_{it-k})$ and apply sample restrictions. The simplest income control is the log of base-year income $ln(z_{i,t-k})$ (Auten and Carroll, 1999). More sophisticated income controls like a spline of $ln(z_{i,t-k})$ are applied as well (Gruber and Saez, 2002). The same is true for sample restrictions. Since mean reversion is pronounced at the bottom of the income distribution, the income distribution is often restricted from below. For instance, typically taxpayers with an income below 10,000$ are excluded from the analysis.

Fourth, variation in marginal tax rates used for identification are assumed to be exogenous. This assumption is violated if tax changes are systematically correlated with other developments that affect economic measures such as GDP. For instance, a tax policy that reduces taxes because policy makers are anticipating a recession is clearly endogenous (Romer and Romer, 2010). If tax changes are correlated with other developments, this leads to a biased estimated of the effect of tax changes. Even previous tax changes can affect current elasticities. Finally, many tax reforms do not target a single income group, and income groups may face different tax rate changes in magnitude throughout the income distribution. To disentangle any non-tax related income changes that systematically vary by income group from the effects of tax rate changes on income becomes even harder, if the extent of tax rate changes is correlated with income.

2.2 Categories of Heterogeneity in Estimated Elasticities

Now, I describe my coded characteristics in more detail. Many factors influence the size of an estimate. To assess the relevance of different explanations, I define various categories of heterogeneity: (1) income concept; (2) estimation techniques; (3) sample restrictions; (4) publication characteristics, including variation across countries and time; and (5) contextual factors. Dimension (1) to (4) are collected from primary studies while dimension (5) is based on external data sources. There are more dimensions of heterogeneity worth investigating, such as the role of income effects, restrictions on demographics (e.g. gender) or tax system-related characteristics (e.g. restricting the sample to individuals who are not eligible for the alternative minimum tax in the US) and even certain control variables such as education. However, a limited number of estimates account for these variables, which makes it impossible test for them. Table 1 provides an overview of all included characteristics and I describe each coded variable in greater detail in the (online) Appendix C.2.

Income Concept. I only distinguish whether or not the dependent variable considers deductions, and I allocate all reported income concepts to two subsamples: before (BD) and after deductions (AD). Kopczuk (2005) shows how the ETI varies with its tax base. While the AD elasticity is considerably larger in a tax system with more deduction possibilities, it can also be lower in a country with a high degree of third party information reporting (e.g. exchange of information between employer and tax authority) (Kleven and Schultz, 2014).

Estimation techniques. I define four distinctive features with respect to estimation techniques that influence the ETI: (a) regression technique; (b) income control; (c) difference length; and (d) weighting by income.

I categorize five *regression techniques*. Since income and marginal tax rates are jointly determined, almost all approaches follow an Instrumental Variable (IV) procedure. They essentially differ in the way they instrument for the net-of-tax rate $(1 - \tau)$. Following Gruber and Saez (2002) the most standard approach is defined as 'IV: mechanical tax rate changes.' The second estimation technique is called 'IV: (lagged) mechanical tax rate changes' (Weber, 2014). Different instruments have recently been developed. For instance, Burns and Ziliak (2017) use a Wald-type grouping instrumental variables estimator. Instead of using a person-specific instrument, they construct a new instrument, which is the cohort-state-year mean of the synthetic tax rate. I summarize all other types of instruments in a third category (IV:other).[8] The earliest method, namely a basic Difference-in-Differences (DID) approach, uses a defined treatment and control group without any instruments and income

[8] All 'other' instruments are explained in greater detail in the (online) Appendix C.2.

controls (Feldstein, 1995). Difference-in-Differences (DID) with a dummy variable as an instrument represents another category. This is a conventional DID approach in which the NTR is instrumented by the interaction of the after-reform and treatment group dummy. This is similar to Feldstein's (1995) tabulated DID approach, but estimated in a regression framework that allows for additional control variables (Moffitt and Wilhelm, 2000).

I define five generations of *income control variables*. First, there is the use of no additional income control variables (none). Studies published prior to 2000 use no income controls and most studies estimate a specification with no income controls as a sensitivity check. The second generation covers studies that use only the log of base-year income control $ln(z_{i,t-k})$ (Auten and Carroll, 1999). Following Gruber and Saez (2002) researchers use more sophisticated income controls like a spline of log base-year income. A spline divides income groups into deciles to account for non linear income trends across these groups. Kopczuk (2005) argues that using only base-year income and some flexible function is not sufficient. He explicitly distinguishes between permanent and transitory income components and proposes two types of income control variables: the log of lag base-year income $ln(z_{i,t-k-1})$, which allows one to control for an individual's rank in the income distribution and therefore for the permanent income level; and transitory income trends are captured by using the deviation between log base-year and log lag base-year income $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$. The last generation covers every *other* (non-standard) income control used in the literature, e.g. cohort-state-year income control as used in Burns and Ziliak (2017).

All studies apply a 'First Difference' estimation strategy with a varying *difference length* to eliminate the impact of unobservable time-invariant characteristics. An estimate is either based on a specification with a time window of 1-year, 2-year, 3-year or of 4 and more years. The chosen difference length $t-k$ has an effect on resulting estimates. Most estimations use a 3-year time window such that researchers relate income and marginal tax rates e.g. from 2001 to 2004. One might think that the longer the time window, the larger the behavioural response. However, the timing, announcement and implementation of underlying reform(s), individual speed of understanding, as well as an individual's ability to adjust their income have an effect on the size of behavioural adjustments. Since many tax reforms are phased-in over several years, an estimate is only a combination of short-, medium- and long-run responses (Weber, 2014).

Since weighted elasticity parameters reflect the relative contribution to total revenues, regression results are sometimes *weighted by income* (Gruber and Saez, 2002).[9] If responses do not vary by income, weighting the estimates by income will not affect elasticity estimates. However, it seems reasonable to assume that behavioural responses are not homogeneous across the income distribution. Weighted results account for the fact that high-income taxpayers tend to exhibit larger responses. Typically, these weights are censored at the top (e.g. at 1 $ million) and are not free of criticism, since income itself is endogenous (Weber, 2014). Individuals who face a temporary positive income shock will receive a larger weight. The weight is even larger if high-income earners are affected. Hence, resulting estimates are even more strongly distorted.

Sample restrictions. I coded whether *income cutoffs* are used and, if so, the corresponding threshold. These thresholds are re-calculated in US-Dollar. To account for mean reversion at the beginning and end of an individual's working life, researchers apply an *age cutoff* to limit the sample to the working population and to exclude pensioners.[10]

Publication characteristics and variations across countries and time. To account for potential differences, I control for whether or not an estimate is reported in a journal or in a working paper. Given the research process, I include different categories for *publication decade* ((1) <= 2000, (2) 2001-2010; (3) >2010) as controls. Publication decade does not necessarily coincide with the timing of a tax reform. To identify a potential development over time which is not directly related to any type of methodological progress but rather related to tax policy at a given time, I include *estimation/ data decade* as a control. For a particular estimate,

[9] Similar to missing details regarding whether or not capital gains are included in the income concept, it often remains unclear by what type of income estimates are weighted.

[10] In the (online) Appendix D, I provide estimation results that account for sample restrictions with respect to *marital status* and *employment type*.

I calculate the mean of the first and end years of the underlying data period ('mean year of observation') and assign the corresponding decade: 1980s, 1990s or 2000s. Countries are summarized in different *country groups* (1) USA, (2) Scandinavia (Denmark, Norway, Sweden), and (3) other countries (Canada, Finland, France, Germany, Hungary, Netherlands, New Zealand, Poland, Spain).

Contextual Variables. Inequality measures and economic characteristics shape behavioural responses to taxation. To account for income inequality within an economy, I include the *Gini coefficient* (disposable income, post taxes and transfers). In addition, I consider a measure of the share of pre-tax national income that is held by the *top 1%* and *top 10%* as contextual variables in my regression. An increase in inequality might be the result of past tax cuts for high-income tax payers. Hence, larger estimates might not be the result of larger responses, but rather of a widening in the income distribution that is captured by estimated elasticities.

Aspects of a given tax system as well as the underlying business cycle are related to behavioural responses to taxation. Kleven and Schultz (2014) find that behavioural elasticities are larger when estimated from large tax reform episodes and a more salient tax reform is more likely to overcome optimization frictions. Therefore, I account for the *introduction of a top tax bracket*. Since such a reform is more salient and the affected tax group is the most responsive one, this might lead to higher estimates.[11] Hargaden (2015) provides evidence of a weaker behavioural response during a recession and therefore highlights the role of business cycle fluctuations. To account for a given economic situation, I add the respective *unemployment rate* as a contextual variable in my regression.

Third party information reporting (e.g. the exchange of information of employers or banks and tax authorities) plays a key role in tax compliance and a country's overall tax take. Kleven et al. (2011) find that the overall tax evasion rate is very small in Scandinavia because

[11] Tax reforms are necessary to generate variation that can be exploited. A reform does not happen in a single year, nor is it easy to tell exactly which income group is affected. Moreover, most estimates are based on a data period with more than one single change in tax law. This makes it difficult to account for other tax reform characteristics in the meta analysis.

almost all income is subject to third party information reporting. I include two variables as a proxy to check for its influence. First, the *fraction of self-employed* workers within a country. Traditionally, self-employed taxpayers provide most of the necessary information to tax authorities themselves. I expect a positive relationship between elasticities and the share of self-employed workers within an economy. As a second measure, I include the share of *modern taxes per GDP* to proxy for the share of tax revenue that are exposed to third-party information reporting compared to the overall tax take. Kleven et al. (2016) distinguish between what they call traditional and modern taxes. Unlike traditional taxes, which rely on self-reported information, modern taxes rely on third-party information.[12] I expect a negative correlation between reported elasticities and modern taxes to GDP ratio.

2.3 Descriptive Statistics

Table 1 provides an overview of the collected information to explain differences in elasticity estimates. As already mentioned, I divide the meta-sample in two subsamples depending on whether the underlying income concept accounts for deductions. The before deductions subsample consists of 940 observations collected from 46 studies and the after deduction subsample of 780 observations from 41 studies.

Around 60% of the estimates refer to a regression technique that uses mechanical tax rate changes as an instrument. One third of estimates use the log of base year income (Auten and Carroll, 1999) as an income control. Most estimates either use a difference length of three years or consider a short time window of one year. 40-50% of all primary estimates are weighted by income. Almost half of the estimates apply an age cutoff and the vast majority of estimates use an income cutoff.

[12]Modern taxes are defined as personal and corporate income taxes, value-added taxes, payroll taxes, and social security contributions, whereas traditional taxes are all other taxes (e.g. inheritance tax). Modern taxes play a crucial role in the economic development of a country and there is a strong positive correlation between GDP per capita and modern taxes to GDP.

Table 1: Descriptive Statistics: Categories of Heterogeneity

	Before Deductions (BD) (N=940)		After Deductions (AD) (N=780)	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Estimation Techniques				
Regression technique				
IV: mechanical tax rate changes	0.651	0.477	0.609	0.488
IV: (lagged) mechanical tax rate changes	0.041	0.20	0.165	0.372
IV: other	0.094	0.291	0.127	0.333
DID and IV	0.188	0.391	0.045	0.207
classic DID	0.026	0.158	0.054	0.226
Income Control				
Auten Carroll (1999)	0.286	0.452	0.226	0.418
none	0.206	0.405	0.224	0.417
Gruber Saez (2002) spline	0.181	0.385	0.176	0.381
Kopczuk (2005) type	0.249	0.433	0.353	0.478
other	0.078	0.268	0.022	0.146
Difference Length				
3 years	0.395	0.489	0.512	0.500
1 year	0.366	0.482	0.287	0.453
2 years	0.074	0.263	0.128	0.335
4+ years	0.165	0.371	0.073	0.260
Weighted by Income	0.484	0.500	0.405	0.491
Sample Restrictions				
Age Cutoff	0.564	0.496	0.523	0.5
Income Cutoff				
0-10k	0.255	0.436	0.236	0.425
none	0.127	0.333	0.199	0.399
10k-12k	0.249	0.433	0.292	0.455
12-31k	0.191	0.394	0.114	0.318
$>$ 31k	0.178	0.382	0.159	0.366
Variations across Countries and Time				
Country Group				
USA	0.494	0.500	0.532	0.499
Scandinavia	0.184	0.388	0.099	0.298
other countries	0.322	0.468	0.369	0.483
Mean year in study data	1994.524	7.819	1995.976	8.849
Estimation decade				
$<$ 1999	0.286	0.452	0.288	0.453
1990 - 2000	0.394	0.489	0.262	0.440
$>$ 2000	0.320	0.467	0.450	0.498
Publication Characteristics				
Publication decade				
2001-2010	0.367	0.482	0.414	0.493
$<=$ 2000	0.063	0.243	0.033	0.180
$>$ 2011	0.570	0.495	0.553	0.498
Published Type				
published	0.671	0.470	0.654	0.476
working paper	0.126	0.331	0.101	0.302
Mean Year of Publication	2011.169	0	2011.169	0
Contextual Variables				
Gini	30.908	5.178	31.684	4.445
top 10% inc. share	0.333	0.059	0.341	0.061
top 1% inc. share	0.109	0.034	0.114	0.037
intro top bracket	0.278	0.448	0.218	0.413
unemployment rate	6.917	2.874	7.023	1.638
fraction self-employed	10.934	3.833	11.056	3.6
share of modern taxes	26.688	9.195	25.449	8.388

Note: I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic. Reference categories are given in italics. More details can be found in the online Appendix C.2. For a given estimate, contextual variables are merged via country and/or mean year of observation.

3 Meta-Regression Results

In Section 3.1, I separately present the results for before (BD) and after deduction (AD) elasticities. In addition, I present some stylized elasticity estimates. These estimates are intended to facilitate the interpretation of my results and to summarize results that correspond to the two most commonly applied approaches in the literature. In Section 3.2, contextual characteristics will be analysed separately and I show that both BD and AD elasticities are correlated with tax system- and economy related characteristics.[13]

3.1 Baseline Results

I run specification (1) on the *before* and *after* deduction subsample separately and present the results in Table 2 and 3. I define the most commonly used characteristic as a reference category (written in bold) and omit this feature such that reported coefficients need to be interpreted as a deviation from a particular characteristic to the corresponding reference category. I gradually add the defined characteristics. In column (1) and (2) I only control for estimation technique, and in column (3) I account for sample restrictions. If 'no restriction' defines the base category, it means that a particular estimate is not restricted with respect to a certain characteristic. For instance, the baseline category for age restriction is 'no restriction.' Hence, estimates need to be interpreted in reference to other estimates that do not apply an age cutoff. Results on country group coefficients are presented in column (4) and (5), with column (4) accounting for (estimation) decade, column (5) controlling for (publication) decade. Column (6) presents the most comprehensive specification.[14] Baseline results do not account for contextual factors. The reference specification in column (1) is defined as a specification that uses mechanical tax rate changes as an instrument, log base-year income control and a three-year difference length. For example, it refers to the most

[13]To verify the robustness of the baseline results, I apply various estimation techniques and further limit the dataset along certain dimensions (see tables 17 and 18 in the (online) Appendix F).

[14]Multicollinearity might be a problem in the regressions resulting in standard errors that are too large. This makes it difficult to isolate the influence of a single variable from overall influence. Therefore, I check if the variance inflation index is below 10 such that the presented results are reliable within every estimation. Except for column (6) in Table 2 and Table 3 this condition holds.

standard approach used by Kleven and Schultz (2014) in their baseline specifications. On average, such a specification yields a BD elasticity of 0.073 and an AD elasticity of 0.445. As expected, estimates that allow for deduction responses mostly reveal a larger constant and, therefore, are statistically more elastic to marginal tax rate changes compared to results obtained based on the before (BD) subsample. Next, I present results obtained for both subsamples by category of heterogeneity.

Estimation techniques. My results show that AD elasticities are more sensitive with respect to different aspects of the underlying estimation technique compared to BD elasticities. Starting with the choice of *income control*, most studies follow Auten and Carroll (1999) and include log base-year income as an explanatory variable. Compared to this approach a regression that does not consider any income control leads to lower and often negative BD elasticities (a fact that is already noted in most primary studies). My results generalize this finding and quantify an average decrease by 0.2 in BD elasticities. This result is quite robust even in the most sophisticated specification in column (6). All other kinds of income control variables (in most cases more sophisticated ones) lower elasticities in both but in particular in the AD subsample. The success of these controls depends on the extent of year-to-year mean reversion and the stability of the underlying income distribution. However, there is a potential risk that they absorb too much identifying variation (see Saez et al., 2012 for a discussion). It is worth highlighting that Kopczuk-type income controls lower AD elasticities (on average) by 0.371 compared to a log base-year income control while other types of income controls (mostly splines) also decrease AD elasticities but at a lower rate.

The results suggest that the chosen *difference length* has different effects on BD and AD elasticities. In the BD subsample, all specifications with a two-year time window have a marginally lower elasticity compared to specifications based on three-year differences, while there are no statistically significant differences in difference length among AD elasticities. It is reasonable to assume that the result represents different responses. Whereas BD estimates mainly reflect labour supply responses that are not easily to adjust, exploiting tax deductions is an easy way to change an individual's income in response to tax rate changes.

There is no statistical significant difference across BD elasticities if they are *weighted by income* or not, whereas weighted AD elasticities are significantly lower compared to unweighted ones. These results are unexpected, in particular the finding for the AD subsample. If high taxpayers exhibit larger behavioural responses, weighting by income should result in higher estimates. As noted in Weber (2014), weighting by income is a controversial model choice, because income itself is endogenous and it further lead to distorted results. Moreover, the results obtained in primary studies are mixed. For instance, Gruber and Saez (2002) find that an unweighted gross elasticity is substantially lower to the weighted elasticity, while a weighted ETI is very similar to the unweighted ETI. Giertz (2010) on the other hand finds that unweighted ETI estimates are smaller than income-weighted estimates.

Sample Restrictions. An *age cutoff* restricts income and employment fluctuations at the beginning and end of a person's working life. Such a cutoff has contrasting effects on elasticities depending on the subsample. Estimates in the BD subsample are lowered when a primary study restricts its data to a certain age, while I observe a positive effect on AD elasticities. *Income cutoffs* have no effect on estimated BD elasticities. This is in stark contrast to findings for the AD subsample where an income cutoff and its value matters greatly. This is an interesting finding since it is unclear whether or not a certain cutoff (and its level) helps or impairs identification.

Variations across time and countries. Column (4) and (5) take into account *country group*. While column (4) controls for *estimation decade*, column (5) shows the results for *publication decade*. In both subsamples (publication) decade has a significantly larger effect on resulting estimates than (estimation (or data)) decade. Estimates published prior to 2001 are always larger than those published at a later date - even when controlling for various aspects of estimation technique. (Estimation) decade only influences BD estimates. For instance, those BD estimates that rely on a dataset that cover the 1980s are always larger than those in later years. Most of the other findings of Tables (2) and (3) discussed before prevail. Column (6) shows the results of the most comprehensive specification that accounts for all the defined categories of heterogeneity. Unfortunately, multicollinearity seems to influence the results to the extent that the precision of some coefficients vanishes.

Table 2: WLS before deductions baseline results

Dependent Variable: Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.060*	0.061*	0.054*	0.025**	0.025*	0.026
	(0.031)	(0.033)	(0.029)	(0.012)	(0.015)	(0.017)
IV-other	0.075	0.076	0.081*	0.074	0.107*	0.094
	(0.056)	(0.053)	(0.044)	(0.054)	(0.056)	(0.062)
DID-IV	0.298***	0.269***	0.224**	0.257***	0.313***	0.247***
	(0.053)	(0.066)	(0.105)	(0.056)	(0.075)	(0.074)
DID-classic	0.332***	0.304***	0.068	0.187***	0.149**	0.091
	(0.059)	(0.072)	(0.132)	(0.063)	(0.065)	(0.068)
Income Control (omitted: Auten Carroll)						
none	-0.213***	-0.216***	-0.212***	-0.209***	-0.207***	-0.209***
	(0.024)	(0.023)	(0.025)	(0.028)	(0.029)	(0.028)
Gruber Saez Spline	-0.020***	-0.015***	-0.021***	-0.013*	-0.016***	-0.013*
	(0.005)	(0.005)	(0.007)	(0.007)	(0.005)	(0.007)
Kopczuk	-0.017**	-0.014**	-0.014**	-0.015**	-0.010**	-0.012*
	(0.007)	(0.007)	(0.005)	(0.007)	(0.005)	(0.006)
other	-0.034**	-0.070*	-0.029**	-0.020*	-0.009	-0.033*
	(0.017)	(0.039)	(0.013)	(0.012)	(0.009)	(0.019)
Difference Length (omitted: 3-years)						
1 year	0.060	0.054	0.033	0.034	0.012	0.003
	(0.063)	(0.057)	(0.045)	(0.050)	(0.040)	(0.032)
2 years	-0.013	-0.013	-0.030*	-0.035***	-0.035***	-0.041**
	(0.021)	(0.021)	(0.016)	(0.011)	(0.008)	(0.016)
4 years and more	0.082*	0.085*	0.068**	0.009	0.026	0.027
	(0.042)	(0.043)	(0.030)	(0.019)	(0.021)	(0.022)
Weighting by Income (omitted: no restriction)						
Weighting by Income applied		-0.041				-0.046
		(0.025)				(0.040)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied			-0.282**		-0.267	-0.259
			(0.122)		(0.174)	(0.168)
Income Cutoff applied (omitted: 0-10k)						
none			0.018		-0.020*	-0.023
			(0.021)		(0.010)	(0.024)
10k-12k			0.024		-0.015**	-0.015
			(0.016)		(0.007)	(0.011)
12k-31k			0.009		0.007	0.014
			(0.007)		(0.008)	(0.011)
>31k			0.021		-0.005	-0.004
			(0.017)		(0.012)	(0.013)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia				-0.135***	0.239*	0.176
				(0.042)	(0.123)	(0.143)
other countries				-0.020	0.343***	0.300**
				(0.051)	(0.126)	(0.127)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.426**	0.388**
					(0.207)	(0.191)
after 2010					-0.205***	-0.130
					(0.073)	(0.104)
(Estimation) Decade (omitted: 1980s)						
1990s				-0.048***		-0.049***
				(0.002)		(0.002)
2000s				-0.031***		-0.060
				(0.010)		(0.037)
Constant	0.073***	0.110***	0.351***	0.239***	0.296***	0.359***
	(0.007)	(0.028)	(0.123)	(0.041)	(0.054)	(0.059)
Observations	940	940	940	940	940	940
Adjusted R^2	0.566	0.575	0.615	0.637	0.655	0.680

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Baseline results do not account for contextual factors. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: WLS after deductions baseline results

Dependent Variable: Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.409***	0.362***	0.420***	0.333***	0.207***	0.141**
	(0.088)	(0.127)	(0.061)	(0.079)	(0.074)	(0.069)
IV-other	-0.265*	-0.253	-0.246**	0.069	0.197	0.009
	(0.145)	(0.155)	(0.118)	(0.275)	(0.218)	(0.193)
DID-IV	-0.590**	-0.615***	-0.702**	-0.379	-0.289	-0.393
	(0.224)	(0.219)	(0.281)	(0.273)	(0.475)	(0.468)
DID-classic	-0.188	-0.200	-0.189	-0.061	-0.178	-0.166
	(0.372)	(0.334)	(0.363)	(0.402)	(0.305)	(0.281)
Income Control (omitted: Auten Carroll)						
none	0.108	0.074	0.045	0.084	-0.249	-0.237
	(0.078)	(0.084)	(0.089)	(0.087)	(0.159)	(0.163)
Gruber Saez Spline	-0.100	-0.007	-0.137**	-0.110	-0.119	-0.048
	(0.068)	(0.029)	(0.068)	(0.069)	(0.088)	(0.119)
Kopczuk	-0.371***	-0.231***	-0.387***	-0.229**	0.025	0.126
	(0.043)	(0.074)	(0.075)	(0.091)	(0.104)	(0.076)
other	-0.195**	-0.134	-0.331**	-0.066	0.048	0.082
	(0.075)	(0.115)	(0.132)	(0.114)	(0.124)	(0.161)
Difference Length (omitted: 3-years)						
1 year	-0.048	-0.079	0.073	-0.066	0.119	0.103
	(0.106)	(0.131)	(0.074)	(0.121)	(0.090)	(0.105)
2 years	0.033	0.045	0.019	-0.058	0.057	0.053
	(0.086)	(0.081)	(0.119)	(0.078)	(0.105)	(0.109)
4 years and more	0.285	0.187	0.182	0.139	-0.362	-0.399*
	(0.191)	(0.200)	(0.212)	(0.204)	(0.242)	(0.235)
Weighting by Income (omitted: no restriction)						
Weighting by Income applied		-0.195**				-0.208
		(0.091)				(0.160)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied			0.252**		0.140	0.187
			(0.113)		(0.124)	(0.138)
Income Cutoff applied (omitted: 0-10k)						
none			0.154***		0.254***	0.337***
			(0.054)		(0.087)	(0.058)
10k-12k			0.109		0.353	0.514**
			(0.090)		(0.236)	(0.224)
12k-31k			0.111*		0.068	0.093*
			(0.063)		(0.059)	(0.050)
>31k			0.468		0.518	0.625**
			(0.424)		(0.353)	(0.306)
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia				-0.111	0.410	0.477*
				(0.089)	(0.305)	(0.279)
other countries				0.237	0.632**	0.608*
				(0.215)	(0.304)	(0.312)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					1.164*	1.203*
					(0.662)	(0.607)
after 2010					-0.500***	-0.417*
					(0.173)	(0.221)
(Estimation) Decade (omitted: 1980s)						
1990s				-0.018		-0.043
				(0.060)		(0.040)
2000s				-0.185		0.030
				(0.242)		(0.131)
Constant	0.445***	0.496***	0.208***	0.420***	-0.019	-0.082
	(0.040)	(0.059)	(0.066)	(0.103)	(0.272)	(0.243)
Observations	780	780	780	780	780	780
Adjusted R^2	0.405	0.423	0.479	0.425	0.621	0.633

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Baseline results do not account for contextual factors. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Summary. To highlight the sensitivity of both types of elasticities with respect to the estimation technique, I calculate some stylized elasticity estimates. In Table 4 I present average BD and AD elasticity estimates for the most commonly used specifications. The upper part of the table considers an approach that uses mechanical tax rate changes as an instrument and a difference length of three years (= (basic) Gruber Saez approach). The lower part considers an approach that uses (lagged) mechanical tax rate changes and a difference length of two years (= (basic) Weber approach). Both parts show the results for various income controls. Compared to BD elasticities the magnitude of AD elasticities is not only larger by definition but AD elasticities are also more sensitive with respect to aspects of the chosen estimation technique. Average BD elasticities lie in the range of 0.053 to 0.120, while average AD elasticities vary from 0.074 to 0.887. Richer (or more sophisticated) income controls always lower elasticities and the effect is more pronounced in the AD subsample.

Table 4: Stylized elasticity estimates

An approach that uses the following characteristics leads to:	BD	AD
IV: mechanical tax rate changes		
Difference Length of 3 years		
and the following income controls:		
Auten Carroll	0.073	0.445
Gruber Saez Spline	0.053	0.345
Kopczuk	0.056	0.074
IV: (lagged) mechanical tax rate changes		
Difference Length of 2 years		
and the following income controls:		
Auten Carroll	0.120	0.887
Gruber Saez Spline	0.100	0.787
Kopczuk	0.103	0.516

Notes: Stylized elasticity estimates are based on results presented in column 1 in tables 2 and 3. For instance, a specification that uses (i) (lagged) mechanical tax rate changes, (ii) a difference length of 2 years and (iii) Kopczuk-type income controls provide an average AD elasticity of 0.516 = 0.445 + 0.409 + 0.033 - 0.371 (compare table 3). Column BD refers to before and column AD to after deduction elasticity estimates.

3.2 Contextual Factors

The following descriptive analysis shows how various contextual factors are associated with the size of elasticity estimates. The baseline specification involves controls for estimation technique, income controls and difference length (see column (1) from Tables 2 and 3). I use this specification and gradually take into account contextual factors as defined in Section 2.2. The exercise shows that past as well as current (tax-) policy and the underlying context matters when interpreting elasticities. The relevant coefficients are displayed in Table 5.

Table 5: WLS: Contextual Factors

Dependent Variable: Elasticity:	Before Deduct.	After Deduct.
Additional Variables		
Gini Coefficient	0.008***	-0.002
	(0.002)	(0.014)
Top 10%	0.814*	3.563**
	(0.442)	(1.536)
Top 1%	0.330	7.709**
	(0.448)	(3.202)
Intro top bracket	-0.026	-0.086
	(0.094)	(0.117)
Unemployment Rates	-0.007	0.067*
	(0.004)	(0.039)
Fraction of self-employed	0.016***	-0.022
	(0.006)	(0.023)
Modern taxes (in 2005)	-0.010***	0.016
	(0.002)	(0.012)

Note: Both columns are estimated using Weighted Least Squares with precision as weights. Standard errors (in parentheses) are clustered at the study level. The baseline specification only includes controls for estimation technique (regression technique, income control and difference length) (same as column (1) from Tables 2 and 3. I gradually add each contextual characteristic separately. For the first characteristic, I compare the first and last year of a data period. Remaining characteristics are merged via mean year of observation. For observations that are based on a classic DID approach, I do not have information of the share of self-employed people that corresponds to the respective mean year of observation. Full results can be found in the (online) Appendix E (see Tables 15 and 16). Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

There is a positive correlation between *inequality measures* and elasticities. In particular, AD elasticities are highly correlated with top income shares. An increase in inequality might be the result of past tax cuts for high-income taxpayers. As Alvaredo et al. (2013) observe, there has been a widening of the income distribution and top tax rates have moved in the opposite direction from top pre-tax income shares. While top pre-tax income shares are rising, top tax rates are decreasing. Such widening in the income distribution affects estimated elasticities. It might be the case that income control variables do not fully account for such a development and this leads to an upward bias of AD elasticities. This confirms the fact that not only current but also past tax policy still has an effect on estimated elasticities and that the underlying context matters when interpreting elasticities.

Given that wealthier people tend to be more responsive, I expect a positive relationship between an *introduction of a top tax bracket* and behavioural responses. Contrary to my expectation, the coefficient is insignificant and close to zero.[15] Business cycle effects are approximated by *unemployment rate* are weakly related to AD elasticities and I do not find any correlation with BD elasticities.

As shown by Kleven et al. (2016), there is a close relationship between tax enforcement, tax compliance and third party information reporting. My regression results show that the share of tax revenue that are exposed to third-party information reporting within a country (*modern taxes per GDP*) is negatively related to BD elasticities. Given that self-employed people have greater control over their income, there is a positive correlation between BD elasticities and the *fraction of self-employed* workers in an economy. Neither measure influences AD elasticities. This strengthens the fact that AD responses are mainly driven by avoidance behaviour. Most taxpayers respond via itemized deductions that are not subject to third party information reporting. The magnitude of estimated elasticities are affected by the degree of third party information reporting which can be influenced by policy makers.

4 Selective Reporting Bias

In the last part of my analysis, I check for the presence of a selective reporting bias. Publishing statistical results that reject the hypothesis of no effect reflects a general desire. Moreover, researchers naturally want to publish results that exhibit intuitive magnitudes.

[15] I ignore all other tax system-related issues (e.g. base broadening) that might have been occurring simultaneously.

Publication or reporting selection bias has been identified in other areas of empirical work. Ashenfelter et al. (1999) review the literature on the rate of return on schooling investment and show reporting selection bias in favour of significant and positive returns to education. Card and Krueger (1995) find such biases in the minimum wage literature and Lichter et al. (2015) in the literature on labour demand elasticities. A study by Brodeur et al. (2016) uses more than 50,000 tests published in three top economic journals and find that researchers are prone to choose more 'significant' specifications in order to increase the chance of publication. Moreover, they show that scientists use z-statistics of 1.64 or 1.96 as reference points.

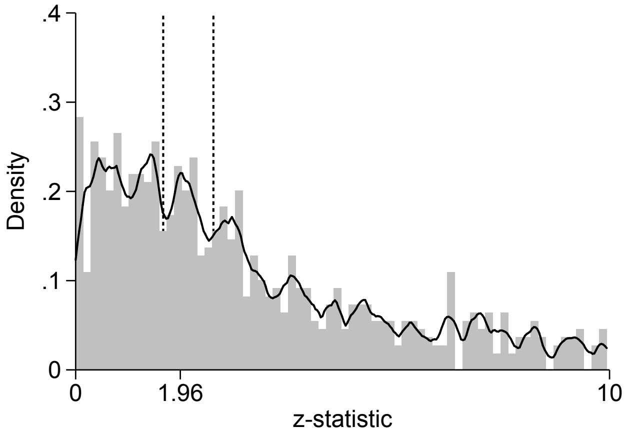
To start the analysis, I follow Brodeur et al. (2016) and plot the distribution of z-statistics and, I then examine the relationship between standard errors and estimates and the distribution of elasticity estimates. Finally, I check statistically whether publication bias is prevalent.

Distribution of z-statistics An obvious type of bias is the excessive production and selection of significant results. Given that z-statistic = beta coefficient/standard error, there are three ways to receive significant values. First, to find a specification where standard errors are low enough. Second, to search for a specification where coefficients are large enough to offset 'large' standard errors. Or third, through a combination of these two things. Since research on behavioural responses to taxation relies on administrative datasets with a large number of observations, standard errors are generally small.

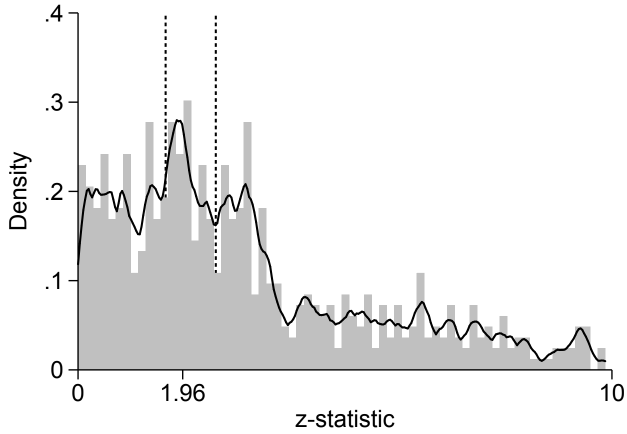
I plot the distribution of z-statistics for the two subsamples (see Figure 2).[16] Subfigure (a) shows the BD and Subfigure (b) the AD subsample. In accordance with Brodeur et al. (2016), I observe a local maximum around 2 (= 5% significance) and also a valley before this. Moreover, I also observe a spike around 1.64 (= 10% significance) and around 3 (= 0.05-0.01%

[16] I formally tested the equality across distributions. I applied a Kolmogorov-Smirnov test which tests whether different t-distributions are equal. More specifically, I test (i) whether the t-statistics of before and after deduction distribution elasticities differ, (ii) within the AD subsample, I check whether the distribution of t-statistics from published estimates and estimates collected from working papers differ, and last (iii) within the BD subsample, I check whether the distribution of t-statistics before and after 2009 differ. In all three cases, I am able to reject the hypothesis that this is the case.

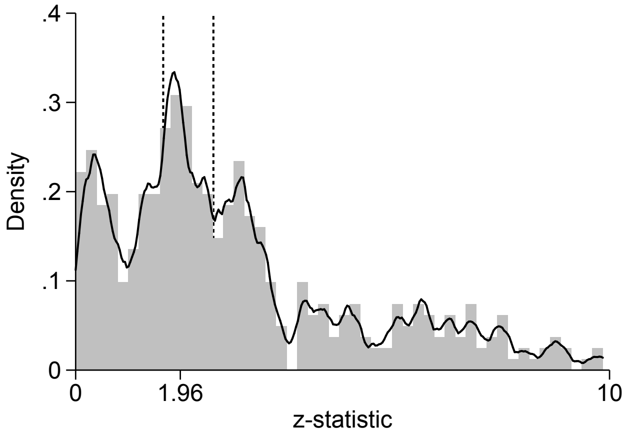
Figure 2: Raw distribution of z-statistics



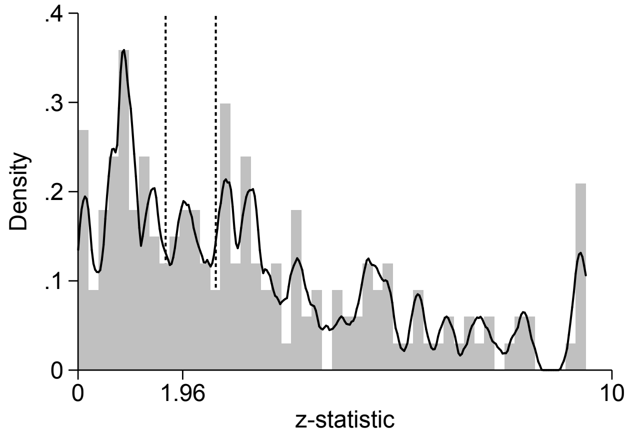
(a) Before Deductions - all



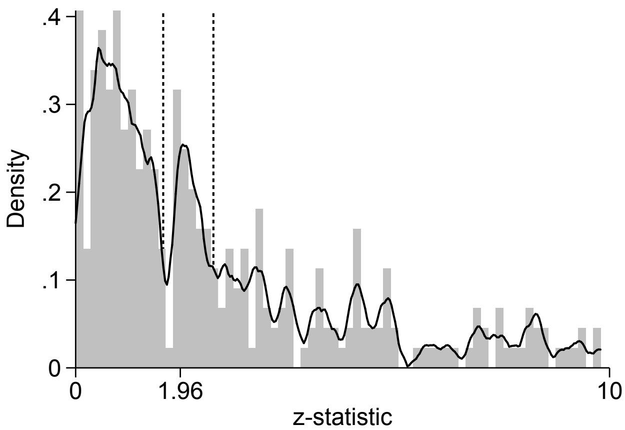
(b) After Deductions - all



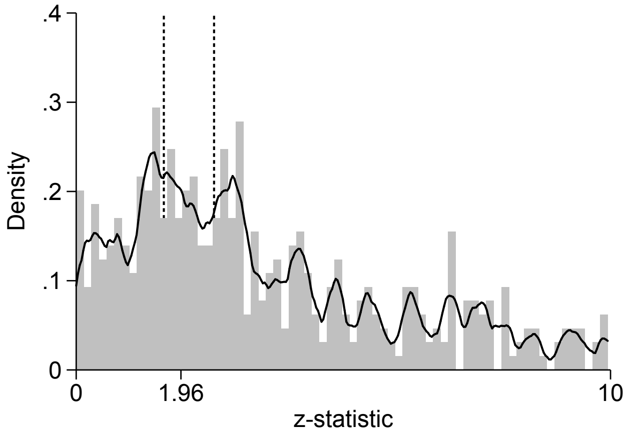
(c) After Deductions - only published



(d) After Deductions - only working paper



(e) Before Deductions - prior to Chetty (2009)



(f) Before Deductions - after Chetty (2009)

Note: All graphs plot the distribution of z-statistics. The significance level of 5% (1.96) and also the z-values for the 10% and 1% level of significance are highlighted. Subfigure (a) plots all estimates from the Before Deductions (BD) subsample and Subfigure (b) for the After Deductions (AD) subsample. Subfigures (c) and (d) split the AD subsample into estimates published in journals and estimates reported in working papers. Subfigures (e) and (f) split the BD subsample into estimates that are published prior to and after 2009.

significance).[17] These simple graphs provide evidence consistent with the existence of 'p-hacking.' This pattern is more pronounced in the AD subsample because researchers usually use the elasticity of taxable income (and not necessarily the elasticity of broad income) when they apply optimal tax rate formulas.

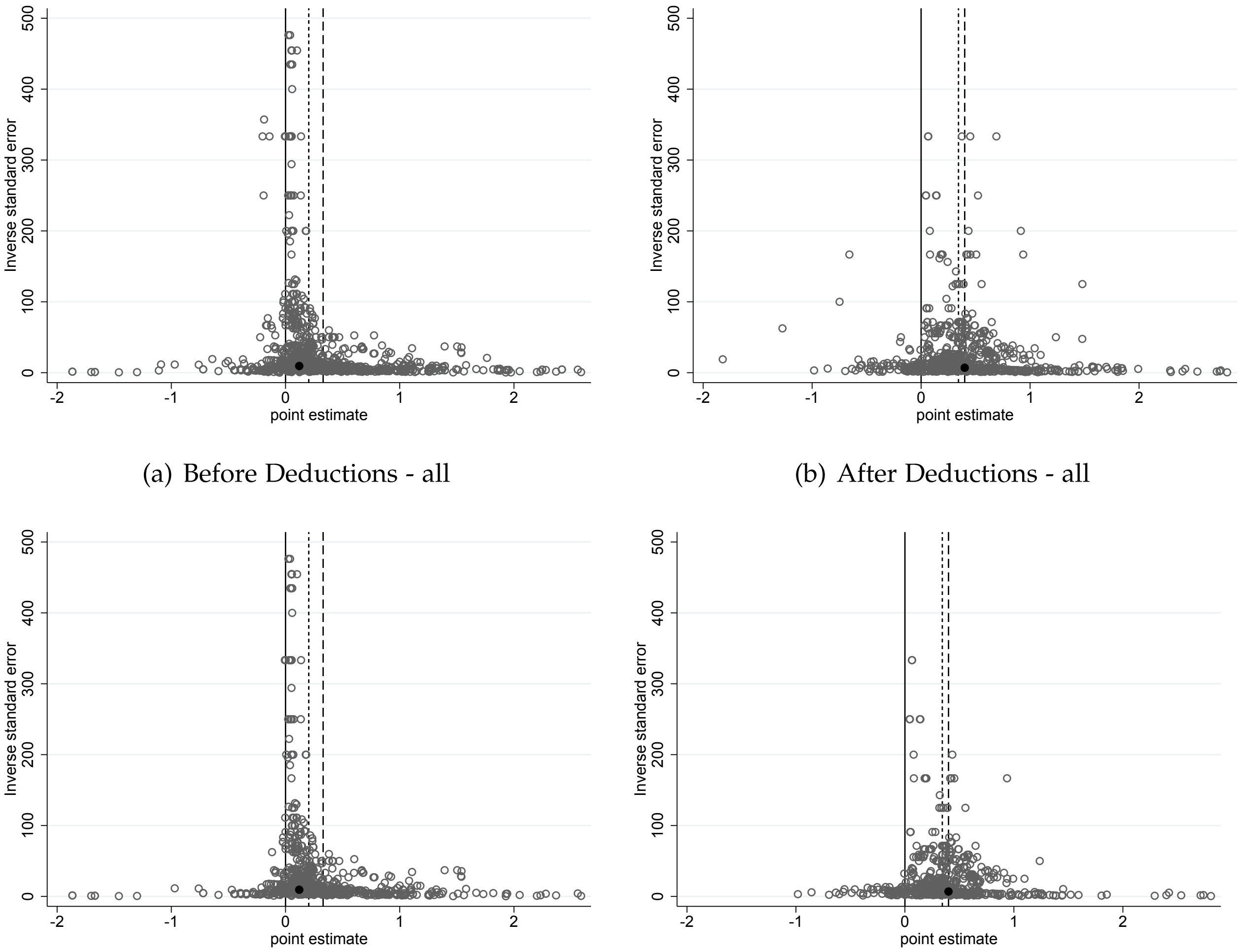
In Subfigure (c) and (d) I divide the AD subsample into estimates reported in journal articles and working papers. The maximum around 2 is even more pronounced for published AD elasticities. It is unclear whether a researcher chooses the most credible findings in the first place to increase the chances of publication and/or that referees/journals prefer significant estimates. Moreover, journal editors often require authors to streamline their papers prior to publication, leading them to limit the number of tables and figures in their paper. Therefore, it is unclear who chooses which estimates are published.

Chetty (2009) shows that the excess burden of taxation depends on a weighted average of taxable income and total earned income elasticities. Since the publication of his study in 2009, BD (e.g. gross income) elasticities have begun to receive more attention. Therefore, I divide the BD subsample into estimates reported prior to and after 2009. As seen in Subfigure (e), I observe a larger insignificant mass before 2009 and a huge spike at 1.96 (=5% significance level) and a missing mass before. After 2009 I observe a much smaller insignificant mass but still a spike at 1.64 (=10% significance), 1.96 (=5% significance) and now also around 3 (=0.05-0.01% significance level). The graphical evidence confirms that the share of significant BD elasticities has increased over time.

Relationship between estimate and standard error. In the second step, I follow Card and Krueger (1995) and analyse the relationship between an estimate and its standard error. I apply a standard procedure and use what is known as a funnel plot in order to analyse the correlation. Funnel plots are simple scatter plots of elasticity estimates on the horizontal axis and their precision (=inverse of standard error) on the vertical axis. The most precise estimates are close to the top of the funnel and as precision decreases, the dispersion of

[17] There are other peaks and valleys across the distributions. Unlike Brodeur et al. (2016) I use considerably fewer observations, with the result that my graphs appear to be more 'bumpy.'

Figure 3: Funnel Plot



(a) Before Deductions - all

(b) After Deductions - all

(c) Before Deductions - only income control(s)

(d) After Deductions - only income control(s)

Note: Funnel plots are presented separately for the before and after deductions subsamples. The short dashed line denotes the median and the dashed line the mean of the corresponding (full) subsample. In the dataset the median (mean) BD elasticity is 0.185 (0.287) and 0.353 (0.403) respectively for an elasticity that considers deductions. The base results from Gruber and Saez (2002) are highlighted in black. They report coefficients of 0.4 with a standard error of 0.144 for the ETI and 0.12 with a standard error of 0.106 for the elasticity of broad (=gross) income. Subfigures (a) and (b) display all collected estimates. Subfigures (c) and (d) are based on a subset of estimates that rely on a specification with income control(s).

estimates increases. The shape of the graph should look like an inverted funnel. In the absence of selective reporting bias, there should be no systematic relationship between estimates and standard errors. All imprecise estimates should have the same probability of being reported. The funnel should be symmetric with the estimates randomly distributed around the population elasticity. If the estimates are correlated with their standard errors, the funnel can take an asymmetric shape. This might happen when researchers select only significant estimates and/or estimates with a certain sign (e.g. omit negative values) such that their results are consistent with theory.[18]

Figure 3 plots BD and AD elasticities separately along with their precision. I highlight the mean and median as well as estimates obtained by Gruber and Saez (2002). Subfigures (a) and (b) are based on the full sample of estimates, while I restrict the sample to estimates that rely on income controls and therefore explicitly account for non-tax related income growth in Subfigure (c) and (d). Subfigures for BD and AD reveal some noticeable differences. First, I observe a more pronounced missing mass on the negative side in the BD compared with the AD subsample. According to theory an increase in the marginal tax rate lowers the net of tax rate, which in turn should reduce taxable income in the simplest case with no income effects or frictions. If a researcher receives a negative value, this translates into a situation where the government can tax income by 100% while the people earn/work even more. Hence, it seems plausible that researchers tend to put more trust in positive results to keep in line with theory. This behaviour causes a positive relationship between standard errors and estimates. AD elasticities allow a wider range of responses and it is also well-documented that running the exact same specification results in a larger AD elasticity compared to an BD elasticity (Gruber and Saez, 2002). The chance of reporting negative values is therefore larger for an elasticity that does not consider deductions. This might explain why I observe a larger missing mass on the negative side in the BD subsample.

Within the AD subsample, it appears that researchers tend to report an estimate between 0 and 0.4 with a higher probability compared to estimates ranging from e.g. 0.4 to 0.8. I expect a negative relationship between standard errors and estimates and therefore a downward bias of AD estimates.

Distribution of estimates. Another kind of selection reporting bias arises, if researchers use well-known results as a reference point and hence are inclined to report only results

[18]As well as a graphical analysis, I formally checked for funnel asymmetry and conducted a so-called Funnel-asymmetry test as proposed by Egger et al. (1997). In all cases, I am able to reject the hypothesis of funnel symmetry. Besides selective reporting bias, there are other reasons why funnel asymmetry could arise (e.g. data irregularities or low methodological quality of some studies).

that are in line with these findings. Piketty and Saez (2013) write in their handbook chapter that an elasticity of 0.25 seems realistic (same as Chetty, 2009), 0.5 is high and 1 is extreme. As seen in Figure 1, there is a general tendency to report results that lie within an interval of 0 and 1. I observe a considerable excess mass between 0.7 and 1. This indicates an aversion to report a value above 1. In their well-known and widely-cited survey, (Saez et al., 2012, p. 42) refer to their estimates and write '[...]. While there are no truly convincing estimates of the long-run elasticity, the best available estimates range from 0.12 to 0.4. [...]' and '[...] 0.25 corresponds to the mid-range of estimates found in the literature. [...]' With regard to the AD-funnel, there is a slight incline to report values between 0 and 0.4 (=mean of AD estimates in the dataset).

Regression results. To statistically examine the presence of selective reporting bias, I take specification of column (1) of Tables 2 and 3 as the baseline specification (=WLS with estimation technique controls) and explicitly control for an estimate's standard error and other publication-related characteristics. Point estimates and respective standard errors should be independent according to random sampling theory (Card and Krueger, 1995; Stanley and Doucouliagos, 2010). For the sake of interpretation, I normalize the standard error.

Overall, my regression results confirm what can already be seen in figures presented before. The funnel plot for BD estimates indicates selective reporting bias towards positive elasticities. This is confirmed in column (1). Published AD estimates suffer more from 'p-hacking' and I statistically show that selective reporting bias is even more pronounced in journals with a high impact factor among AD elasticities (see column (6)).[19] To account for the fact that larger datasets increase the change of yielding standard errors that are small enough to produce significant and trustworthy results, I calculate the median of observations for each subsample and create a dummy variable if an estimate is based on a dataset that is smaller or larger compared to the median sample size of all other collected estimates. For

[19] I downloaded the IDEAS RePEc simple impact factor (22.06.2016) and working papers receive a value of 0.

BD elasticities, the relationship is significantly positive (see column (3)). In columns (4) and (8) I include a dummy variable indicating if an estimate was reported prior to Chetty (2009). Both aspects influence BD but not AD elasticities.

Table 6: Testing for Selective Reporting Bias

Dependent Variable:	BD	BD	BD	BD	AD	AD	AD	AD
Elasticity:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Standard Error	3.654***	4.084***	0.972	0.652	-0.030	-0.834***	-0.223	-0.360
	(0.719)	(0.845)	(0.812)	(0.988)	(0.203)	(0.294)	(0.354)	(0.530)
Journal impact factor		-0.012				0.030**		
		(0.008)				(0.014)		
Std.Error* Impact Factor		-0.051				0.084***		
		(0.035)				(0.022)		
Dummy if obs > median(obs)			0.771***				-0.066	
			(0.279)				(0.285)	
Std.Error*D if obs > median(obs)			4.375***				0.113	
			(1.142)				(0.540)	
Dummy reported prior to 2009				0.575**				-0.1122
				(0.267)				(0.304)
Std.Error*D reported prior to 2009				3.726***				0.217
				(1.322)				(0.614)
Constant	0.876***	0.982***	0.477***	0.460**	0.424**	-0.027	0.400**	0.416*
	(0.180)	(0.213)	(0.138)	(0.181)	(0.158)	(0.221)	(0.158)	(0.248)
Observations	940	940	940	940	780	780	780	780
Adjusted R^2	0.614	0.624	0.628	0.627	0.404	0.456	0.408	0.420

Note: Columns (1) to (8) are estimated using Weighted Least Squares using precision as weights. I control for estimation technique (= regression technique, income control and difference length. Full results can be found in the (online) Appendix G in Tables 19 and 20. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Included standard errors as explanatory variables are normalized such that they can be interpreted as a standard deviation.

Summary. The graphical evidence and regression results indicate an upward reporting bias among BD elasticities, while the reporting bias for AD elasticities goes in both directions with a downward bias appearing to be dominate. The distribution of elasticities an the funnel plot show that there is a tendency to report results that lie within an interval of 0 and 1. In general, reference points related to statistical significance such as 1.96 matters for both types of elasticities and well-known results are targeted. In particular, I observe a larger missing mass for negative values in the BD subsample and I find that researchers report AD estimates ranging from 0 to 0.4 more often compared to results that are located e.g. within 0.4 to 0.8. Among the AD subsample selective reporting bias is even more prevalent in journals with a high impact factor, while the year of publication matters for BD elasticities. Since 2009 have become more significant because of its increased interest.

5 Conclusion

This study applies meta-techniques to identify and to assess different explanations for the varying sizes of estimated elasticities. The magnitude of such estimates is of major importance for tax policy analysis. I differentiate between real responses (before deduction elasticities) and avoidance behaviour (after deduction elasticities) and use 1,720 estimates from 61 studies.

The paper consists of three parts. First, I conduct a meta-regression analysis and quantify the impact of various model choices. Compared to BD elasticities the magnitude of AD elasticities is not only larger by definition, but AD elasticities are also more sensitive with respect to the estimation technique. Second, my study points to correlations between reported estimates and tax system- and economy related characteristics, as well as inequality measures. Last, it shows that selective reporting bias is prevalent in the literature of taxable income elasticities. There is an upward reporting bias among BD elasticities while the reporting bias for AD elasticities goes in both directions with a downward bias appearing to be dominate.

Several important conclusions can be drawn from this analysis. As already acknowledged in the literature, the ETI is not a structural parameter and this study shows that policy conclusions can be misleading. Reported estimates need to be interpreted within the context they are estimated in and researchers and policy makers need to be careful about what type and size of elasticity should be used for policy analysis (e.g. when calibrating an optimal tax model). An application of a simple formula to derive optimal revenue maximising top tax rates, lead to tax rate of 62.5% if I incorporate the mean AD elasticity of 0.403 found in the empirical literature. Using my derived stylized AD estimates ranging from 0.074 to 0.827, lead to tax rates between 44.63% to 90.01%.[20] To develop new (empirical) strategies that are robust to certain model choices, we need to raise the awareness that insignificant and

[20] Assume that the shape of the income distribution in the highest tax bracket is characterised by the Pareto parameter a and e is the elasticity of taxable income or the range of AD elasticities found in this study. Following Saez (2001) the revenue-maximising tax rate is defined as $t = \frac{1}{1+a*e}$. For instance, if $a = 1.5$ and $e = 0.074$, the resulting tax rate is equal to $t = \frac{1}{1+1.5*0.074}$ = 90.01%.

even implausible estimates are meaningful. Instead of proving a single estimate, a range of estimates might help to shed light into the heterogeneity of behavioural responses across the income distribution and different socio-economic groups.

Finally, the literature on taxable income elasticities suffer from selective reporting bias. Unlike the literature on the effects of taxation on labour supply, which relies mostly on survey data, the ETI-literature predominately uses administrative tax-return data.[21] On the one hand, administrative tax-return data provides precise information about a tax unit's income situation that is needed for estimation but, on the other hand, a replication of existing findings is very difficult. Data access is often restricted to a small number of people and its utilisation is costly in various dimensions (e.g. lack of institutional knowledge and language barriers). Future researchers should be encouraged to provide as much information as possible to promote a comprehensive understanding of the obtained elasticities (Slemrod, 2016). Reporting standards or even a pre-analysis plan might reduce the problem of selective reporting bias (see Burlig, 2018; Christensen and Miguel, 2018).

[21] There are some exceptions who either use survey or aggregated administrative data. Recently, Burns and Ziliak (2017) use the Current Population Survey for the US and find elasticities in the range of 0.4 and 0.55. Although deductions and exemptions are precisely measured in administrative tax records, survey data offers a larger set of demographic characteristics and information about the low end of the income distribution. Tax units who do not file a tax return are not available in the tax data and these tax units are in most cases poor households. Future work might consider survey data to (at least) estimate BD elasticities. Saez (2017) provides evidence that even simple tabulated tax data can provide valuable evidence and he points out to possible advantages of such data compared to microlevel data (e.g. simplicity and transparency).

References

Alvaredo, F., Atkinson, A. B., Piketty, T. and Saez, E. (2013), 'The top 1 percent in international and historical perspective', *Journal of Economic perspectives* **27**(3), 3–20.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999), 'A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias', *Labour Economics* **6**(4), 453–470.

Auten, G. and Carroll, R. (1999), 'The Effect of Income Taxes on Household Income', *Review of Economics and Statistics* **81**(4), 681–693.

Bakos, P., Benczúr, P. and Benedek, D. (2010), 'The elasticity of taxable income: Estimates and flat tax predictions using the hungarian tax changes in 2005', *Working Paper* .

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

Burlig, F. (2018), 'Improving transparency in observational social science research: A pre-analysis plan approach', *Economics Letters* **168**, 56 – 60.

Burns, S. K. and Ziliak, J. P. (2017), 'Identifying the elasticity of taxable income', *The Economic Journal* **127**, 297–329.

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

Card, D., Kluve, J. and Weber, A. (2018), 'What works? a meta analysis of recent active labor market program evaluations', *Journal of the European Economic Association* **16**(3), 894–931.

Card, D. and Krueger, A. B. (1995), 'Time-series minimum-wage studies: a meta-analysis', *American Economic Review* **85**(2), 238–243.

Chetty, R. (2009), 'Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss?

The Implications of Evasion and Avoidance', *American Economic Journal: Economic Policy* **1**(2), 31–52.

Christensen, G. and Miguel, E. (2018), 'Transparency, reproducibility, and the credibility of economics research', *Journal of Economic Literature* **56**(3), 920–80.

Diamond, P. A. (1998), 'Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates', *American Economic Review* **88**(1), 83–95.

Egger, M., Smith, G. D., Schneider, M. and Minder, C. (1997), 'Bias in meta-analysis detected by a simple, graphical test', *Bmj* **315**(7109), 629–634.

Feldstein, M. (1995), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act', *Journal of Political Economy* **103**(3), 551–572.

Feldstein, M. (1999), 'Tax Avoidance and the Deadweight Loss of the Income Tax', *Review of Economics and Statistics* **81**(4), 674–680.

Giertz, S. H. (2010), 'Panel Data Techniques and the Elasticity of Taxable Income', *National Tax Association 2009 Annual Conference Proceedings* .

Gruber, J. and Saez, E. (2002), 'The Elasticity of Taxable Income: Evidence and Implications', *Journal of Public Economics* **84**(1), 1–32.

Hargaden, E. P. (2015), 'Taxpayer responses over the cycle: Evidence from irish notches', *Working Paper* .

Harju, J. and Matikka, T. (2016), 'The elasticity of taxable income and income-shifting: what is "real"and what is not?', *International Tax and Public Finance* **23**(4), 640–669.

Havránek, T. (2015), 'Measuring intertemporal substitution: The importance of method choices and selective reporting', *Journal of the European Economic Association* **13**(6), 1180–1204.

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S. and Saez, E. (2011), 'Unwilling or Unable to Cheat? Evidence from a Randomized Tax Audit Experiment in Denmark', *Econometrica* **79**(3), 651–692.

Kleven, H. J., Kreiner, C. T. and Saez, E. (2016), 'Why can modern governments tax so much? an agency model of firms as fiscal intermediaries', *Economica* **83**(330), 219–246.

Kleven, H. J. and Schultz, E. A. (2014), 'Estimating Taxable Income Responses Using Danish Tax Reforms', *American Economic Journal: Economic Policy* **6**(4), 271–301.

Kopczuk, W. (2005), 'Tax Bases, Tax Rates and the Elasticity of Reported Income', *Journal of Public Economics* **89**(11), 2093–2119.

Kreiner, C. T., Leth-Petersen, S. and Skov, P. E. (2016), 'Tax Reforms and Intertemporal Shifting of Wage Income: Evidence from Danish Monthly Payroll Records', *American Economic Journal: Economic Policy* **8**(3), 233–57.

Le Maire, D. and Schjerning, B. (2013), 'Tax bunching, income shifting and self-employment', *Journal of Public Economics* **107**, 1–18.

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

Mertens, K. and Montiel Olea, J. L. (2018), 'Marginal Tax Rates and Income: New Time Series Evidence', *The Quarterly Journal of Economics* **133**(4), 1803–1884.

Mirrlees, J. A. (1971), 'An Exploration in the Theory of Optimum Income Taxation', *The Review of Economic Studies* **38**(2), 175–208.

Moffitt, R. and Wilhelm, M. O. (2000), 'Taxation and the labor supply decisions of the affluent', *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich* p. 193.

Piketty, T. and Saez, E. (2013), 'Optimal Labor Income Taxation', *Handbook of Public Economics* **5**, 391–474.

Romer, C. D. and Romer, D. H. (2010), 'The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks', *American Economic Review* **100**(3), 763–801.

Rosenthal, R. (1979), 'The File Drawer Problem and Tolerance for Null Results.', *Psychological Bulletin* **86**(3), 638.

Saez, E. (2001), 'Using Elasticities to Derive Optimal Income Tax Rates', *Review of Economic Studies* **68**(1), 205–229.

Saez, E. (2010), 'Do Taxpayers Bunch at Kink Points?', *American Economic Journal: Economic Policy* **2**(3), 180–212.

Saez, E. (2017), 'Taxing the rich more: Preliminary evidence from the 2013 tax increase', *Tax Policy and the Economy* **31**(1), 71–120.

Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The Elasticity of Taxable Incomewith Respect to Marginal Tax Rates: A Critical Review', *Journal of Economic Literature* **50**(1), 3–50.

Slemrod, J. (2016), 'Caveats to the Research Use of Tax-Return Administrative Data', *National Tax Journal* **69**(4), 1003.

Slemrod, J. and Kopczuk, W. (2002), 'The Optimal Elasticity of Taxable Income', *Journal of Public Economics* **84**(1), 91–112.

Stanley, T., Doucouliagos, C., Giles, M., Heckemeyer, J., Johnston, R., Laroche, P., Nelson, J., Paldam, M., Poot, J., Pugh, G. et al. (2013), 'Meta Analysis of Economics Research Reporting Guidelines', *Journal of Economic Surveys* **27**(2), 390–394.

Stanley, T. and Doucouliagos, H. (2010), 'Picture this: a simple graph that reveals much ado about research', *Journal of Economic Surveys* **24**(1), 170–191.

Weber, C. (2014), 'Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences', *Journal of Public Economics* **117**, 90–103.

Online Appendices

A Meta (Estimation) Sample

A.1 List of Included Studies

Aarbu, K. O. and Thoresen, T. O. (2001), 'Income Responses to Tax Changes – Evidence from the Norwegian Tax Reform', *National Tax Journal* 54(2), 319–335.

Almunia, M., and Lopez-Rodriguez, D. (2019), 'The Elasticity of Taxable Income in Spain: 1999-2014', *Working Paper (July 2019) forthcoming: SERIEs - Journal of the Spanish Economic Association*.

Arrazola, M., de Hevia, J., Romero, D. and Sanz-Sanz, J. F. (2014), 'Personal Income Tax Reforms and the Elasticity of Reported Income to Marginal Tax Rates: An Empirical Analysis Applied to Spain', *Working Paper Series in Public Finance provided by Victoria Business School (September 2014)*.

Arrazola-Vacas, M., Sanz-Sanz, J. F., , Rueda-Lopez, N. and Romero-Jordan, D. (2015), 'Reported Gross Income and Marginal Tax Rates: Estimation of the Behavioural Reactions of Spanish Taxpayers', *Applied Economics* 47(5), 466–484.

Auten, G. and Carroll, R. (1999), 'The Effect of Income Taxes on Household Income', *Review of Economics and Statistics* 81(4), 681–693.

Auten, G., Carroll, R. and Gee, G. (2008), 'The 2001 and 2003 Tax Rate Reductions: An Overview and Estimate of the Taxable Income Response', *National Tax Journal* 61(3), 345–364.

Auten, G. and Joulfaian, D. (2009), 'The Taxable Income Elasticity of High-Income Taxpayers: Evidence from a Long Panel', *Working Paper (May 2009)*.

Auten, G. and Kawano, L. (2014), 'How the Rich Respond to Anticipated Tax Increases: Evidence from the 1993 Tax Act', *Working Paper (May 2014)*.

Bakos, P., Benczúr, P. and Benedek, D. (2010), 'The Elasticity of Taxable Income: Estimates and Flat Tax Predictions using the Hungarian Tax Changes in 2005', *Working Paper (September 2010)*.

Berg, K. and Thoresen, T. (2018), 'Problematic Response Margins in the Estimation of the Elasticity of Taxable Income' , *Working Paper (June 2018)*.

Blomquist, S. and H. Selin (2010), 'Hourly wage rate and taxable labor income responsiveness to changes in marginal tax rates', *Journal of Public Economics* 94 (11-12), 878-889.

Bosch N. and de Boer, H.-W. (2019) 'Income and Occupational Choice Responses of the Self-employed to Tax Rate Changes: Heterogeneity across Reforms and Income' , *Labour Economics* 58, 1-20.

Burns, S. K. and Ziliak, J. P. (2017), 'Identifying the elasticity of taxable income', *The Economic Journal* 127, 297–329.

Carey, S., Creedy, J., Gemmell, N. and Teng, J. (2015), 'Estimating the Elasticity of Taxable Income in New Zealand', *Economic Record* 91(292), 54–78.

Carroll, R. (1998), 'Do Taxpayers Really Respond to Changes in Tax Rates? Evidence from the 1993 Tax Act', *Office of Tax Analysis Working Paper* (79), 145–77 (November 1998).

Chetty, R., Friedman, J. N., Olsen, T. and Pistaferri, L. (2011), 'Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records', *Quarterly Journal of Economics* 126, 749–804.

Creedy J., Gemell, N. and Teng J. (2018) 'Income Effects and the Elasticity of Taxable Income' , *New Zealand Economic Papers* 52(2), 158-203.

Díaz-Caro, C. and Onrubia, J. (2018) 'How do Taxable Income Responses to Marginal Tax Rates Differ by Sex, Marital Status and Age? Evidence from Spanish Dual Income Tax' , *Economics: The Open-Access, Open-Assessment E-Journal* 2018-67(12), 1-25.

Doerrenberg, P., Peichl, A. and Siegloch, S. (2017), 'The Elasticity of Taxable Income in the Presence of Deduction Possibilities', *Journal of Public Economics* 151, 41–55.

Ericson, P., Flood, L. and Islam, N. (2015), 'Taxes, Wages and Working Hours', *Empirical Economics* 49(2), 503–535.

Gelber, A. M. (2014), 'Taxation and the Earnings of Husbands and Wives: Evidence from Sweden', *Review of Economics and Statistics* 96(2), 287–305.

Giertz, S. H. (2007), 'The Elasticity of Taxable Income over the 1980s and 1990s', *National Tax Journal* pp. 743–768.

Giertz, S.H. (2010), 'Panel Data Techniques and the Elasticity of Taxable Income', *National Tax Association 2009 Annual Conference Proceedings*

Giertz, S. H. (2010), 'The Elasticity of Taxable Income during the 1990s: New Estimates and Sensitivity Analyses', *Southern Economic Journal* 77(2), 406–433.

Gottfried, P. and Schellhorn, H. (2004), 'Empirical Evidence on the Effects of Marginal Tax Rates on Income – The German Case', *IAW-Diskussionspapiere* (15) (February 2004).

Gottfried, P. and Witczak, D. (2009), 'The Responses of Taxable Income Induced by Tax Cuts - Empirical Evidence from the German Taxpayer Panel', *IAW-Diskussionspapiere* (57) (November 2009).

Gruber, J. and Saez, E. (2002), ‘The Elasticity of Taxable Income: Evidence and Implications‘, *Journal of Public Economics*, vol. 84(1), pp. 1–32.

Hansson, Å. (2007), ‘Taxpayers’ Responsiveness to Tax Rate Changes and Implications for the Cost of Taxation in Sweden’, *International Tax and Public Finance* 14(5), 563–582.

Harju, J. and Matikka, T. (2016), ‘The Elasticity of Taxable Income and Income-Shifting: What is “real“and what is not?’, *International Tax and Public Finance* 23(4), 640–669.

Heim, B. T. (2010), ‘The Responsiveness of Self-Employment Income to Tax Rate Changes’, *Labour Economics* 17(6), 940–950.

Heim, B. T. and Mortenson, J. A. (2016), ‘The Effect of Recent Tax Changes on Taxable Income: Correction and Update’, *Journal of Policy Analysis and Management* 37(3), 686-694.

He, D., Peng, L. and Wang, X. (2018) ‘Understanding the Elasticity of Taxable Income: A Tale of Two Approaches’ , *Working Paper (September 2018)*

Hermle, J. and Peichl, A. (2018) ‘Jointly Optimal Taxes for Different Types of Income’ , *Working Paper (September 2018)*

Holmlund, B. and Söderström, M. (2011), ‘Estimating Dynamic Income Responses to Tax Reform’, *The BE Journal of Economic Analysis & Policy* 11(1), 1–38.

Igdalov, S., Frish, R. and Zussman, N. (2017) ‘The Wage Response to a Reduction in Income Tax Rates: The 2003–2009 Tax Reform in Israel’ , *Working Paper (December 2017)*

Jongen, E. L. and Stoel, M. (2019), ‘The Elasticity of Taxable Labour Income in the Netherlands’, *Working Paper (January 2019)*.

Kemp J. (2017), ‘The Elasticity of Taxable Income: The case of South Africa’, *Working Paper (May 2017)*.

Kiss, A. and Mosberger, P. (2014), ‘The Elasticity of Taxable Income of High Earners: Evidence from Hungary’, *Empirical Economics* 48(2), 883–908.

Kleven, H. J. and Schultz, E. A. (2014), ‘Estimating Taxable Income Responses Using Danish Tax Reforms’, *American Economic Journal: Economic Policy* 6(4), 271–301.

Kopczuk, W. (2005), ‘Tax Bases, Tax Rates and the Elasticity of Reported Income’, *Journal of Public Economics* 89(11), 2093–2119.

Kopczuk, W. (2015), ‘Polish Business Flat Tax and its Effect on Reported Incomes: Preliminary Analysis’, *Working Paper (June 2015)*.

Kumar, A. and Liang, C.-Y. (2017), ‘Estimating Taxable Income Responses with Elasticity Heterogeneity’, *Working Paper (March 2017)*.

Lehmann, E., Marical, F. and Rioux, L. (2013), 'Labor Income Responds Differently to Income-Tax and Payroll-Tax Reforms', *Journal of Public Economics* 99, 66–84.

Lindsey, L. B. (1987), 'Individual Taxpayer Response to Tax Cuts: 1982–1984: With Implications for the Revenue Maximizing Tax Rate', *Journal of Public Economics* 33(2), 173–206.

Looney, A. and Singhal, M. (2006), 'The Effect of Anticipated Tax Changes on Intertemporal Labor Supply and the Realization of Taxable Income', *Working Paper (July 2006)*.

Massarrat Mashhadi, N. and Werdt, C. (2012), 'Estimating Dynamic Income Responses to Tax Changes', *Working Paper (December 2012)*.

Matikka, T. (2016), 'The Elasticity of Taxable Income: Evidence from Changes in Municipal Income Tax Rates in Finland', *Scandinavian Journal of Economics* 120(3), 943-973 .

Miyazaki, T. and Ishida, R. (2016), 'Estimating the Elasticity of Taxable Income: Evidence from Top Japanese Taxpayers', *Working Paper (October 2016)*.

Moffitt, R. and Wilhelm, M. O. (2000), 'Taxation and the Labor Supply Decisions of the Affluent', *Does Atlas Shrug?: The Economic Consequences of Taxing the Rich* p. 193.

Mortenson, J. A. (2016), 'All Income is Not Created Equal: Cross-Tax Elasticity Estimates in the United States', *Working Paper (March 2016)*.

Pirttilä, J. and Selin, H. (2011a), 'Income Shifting within a Dual Income Tax System: Evidence from the Finnish Tax Reform of 1993', *The Scandinavian Journal of Economics* 113 (1), 120–144.

Saez, E. (2003), 'The Effect of Marginal Tax Rates on Income: a Panel Study of 'Bracket Creep', *Journal of Public Economics* 87(5), 1231–1258.

Saez, E., Slemrod, J. and Giertz, S. H. (2012), 'The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review', *Journal of Economic Literature* 50(1), 3–50.

Schmidt, T.-P. and Müller, H. (2012), 'Die Elastizität des zu versteuernden Einkommens in Deutschland: Eine empirische Untersuchung auf Basis des deutschen Taxpayer-Panels', *Working Paper (June 2012)*.

Sillamaa, M.-A. and Veall, M. R. (2001), 'The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1988 Tax Flattening in Canada', *Journal of Public Economics* 80(3), 341–356.

Singleton, P. (2011), 'The Effect Of Taxes On Taxable Earnings: Evidence From The 2001 And Related U.S. Federal Tax Acts', *National Tax Journal* 64(2), 323.

Thomas, A. (2012), 'The Elasticity of Taxable Income in New Zealand: Evidence from the 1986 Tax Reform', *New Zealand Economic Papers* 46(2), 159–167.

Thoresen, T. O. and Vattø, T. E. (2015), 'Validation of the Discrete Choice Labor Supply Model by Methods of the New Tax Responsiveness Literature', *Labour Economics* 37, 38–53

Weber, C. (2014a), 'Measuring Treatment for Tax Policy Analysis', *Working Paper (November 2014a)*.

Weber, C. (2014b), 'Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences', *Journal of Public Economics* 117, 90–103.

Werdt, C. (2015), 'The Elasticity of Taxable Income for Germany and its Sensitivity to the Appropriate Model', *Working Paper (Jan 2015)*.

A.2 Distribution of Estimates by Study: Published vs Working Paper

Published articles Study	# estimates	in %	Working Papers Study	# estimates	in %
Aarbu and Thoresen (2001)	8	0.47	Almunia and Lopez-Rodriguez (2019)	83	4.83
Arrazola-Vacas et al. (2015)	27	1.57	Arrazola et al. (2014)	8	0.47
Auten and Carroll (1999)	20	1.16	Auten and Joulfaian (2009)	24	1.40
Auten et al. (2008)	10	0.58	Auten and Kawano (2014)	12	0.70
Blomquist and Selin (2010)	10	0.58	Bakos et al. (2010)	21	1.22
Bosch (2019)	44	2.56	Berg and Thoresen (2018)	4	0.23
Burns and Ziliak (2017)	68	3.95	Carroll (1998)	12	0.70
Carey et al. (2015)	6	0.35	Giertz (2010)	72	4.19
Chetty et al. (2011)	6	0.35	Gottfried and Schellhorn (2004)	11	0.64
Creedy et al. (2018)	3	0.17	Gottfried and Witczak (2009)	15	0.87
Diaz-Caro and Onrubia (2018)	29	1.69	He et al. (2018)	4	0.23
Dörrenberg et al. (2017)	16	0.93	Hermle and Peichl (2018)	4	0.23
Ericson et al. (2015)	5	0.29	Igdalov et al. (2017)	19	1.10
Gelber (2014)	16	0.93	Jongen and Stoel (2019)	99	5.76
Giertz (2007)	69	4.01	Kemp (2017)	18	1.05
Giertz (2010)	127	7.38	Kopczuk (2015)	30	1.74
Gruber and Saez (2002)	35	2.03	Kumar and Liang (2017)	21	1.22
Hansson (2007)	30	1.74	Looney and Singhal (2017)	15	0.87
Harju and Matikka (2016)	14	0.81	Massarrat Mashhadi and Werdt (2012)	9	0.52
Heim (2010)	14	0.81	Miyazaki and Ishida (2016)	8	0.47
Heim and Mortenseon (2018)	14	0.81	Mortenson (2016)	42	2.44
Holmlund and Söderström (2011)	36	2.09	Schmidt and Müller (2012)	18	1.05
Kiss and Mosberger (2014)	15	0.87	Weber (2014)	5	0.29
Kleven and Schultz (2014)	114	6.63	Werdt (2015)	11	0.64
Kopczuk (2005)	91	5.29			
Lehmann et al. (2013)	18	1.05			
Lindsey (1987)	14	0.81			
Matikka (2018)	18	1.05			
Moffitt and Wilhelm (2000)	39	2.27			
Pirttilä and Selin (2011)	10	0.58			
Saez (2003)	91	5.29			
Saez et al. (2012)	24	1.40			
Sillamaa and Veall (2001)	25	1.45			
Singleton (2011)	25	1.45			
Thomas (2012)	8	0.47			
Thoresen and Vatto (2015)	21	1.22			
Weber (2014)	35	2.03			
Total (published)	1141	67.15%		579	32.85%

Note: The data covers only observations with a given or calculable standard error. # estimates denote the number of estimates collected in a particular study and the corresponding percentage share shows the share a study has in the final sample.

B Additional Descriptives

B.1 Summary Statistics by Income Concept

Table 8: Distributions of Estimates by Income Concept

Tax Base	Mean	Median	Std. Dev.	Obs.	Studies
Before Deductions	0.287	0.185	1.212	940	46
Adjusted Gross Income	0.319	0.236	2.607	278	
Gross Income	0.312	0.230	0.542	414	
Earned Income	0.125	0.062	0.257	129	
Self employed Income	0.675	0.858	0.510	20	
Wage Income	0.230	0.114	0.744	99	
After Deductions	0.403	0.353	0.564	780	41
Taxable Income	0.4	0.343	0.578	737	
Taxable Earnings	0.445	0.444	0.186	43	
Total	0.34	0.270	0.975	1720	61

Note: The data covers only observations with a given or calculable standard error.

B.2 Distribution of Estimates by Country and Income Concepts

Table 9: Income Concepts by Country

Variable	Adj. G. Income	Gross Income	Taxable Income	Earned Income	Self employed	Wage Income	Taxable Earnings	Total
Canada	15	2	2	2	2	2	0	25
China	0	0	4	0	0	0	0	4
Denmark	0	18	18	78	0	6	0	120
Finland	0	6	17	0	0	19	0	42
France	0	0	0	0	0	18	0	18
Germany	3	20	61	0	0	0	0	84
Hungary	0	0	36	0	0	0	0	36
Israel	0	19	0	0	0	0	0	19
Japan	0	0	8	0	0	0	0	8
Netherlands	99	0	44	0	0	0	0	143
New Zealand	0	0	13	0	0	4	0	17
Norway	0	0	12	21	0	0	0	33
Poland	0	30	0	0	0	0	0	30
South Africa	0	9	9	0	0	0	0	18
Spain	0	53	94	0	0	0	0	147
Sweden	12	26	17	12	0	0	30	97
USA	149	231	402	16	18	50	13	879
Total	278	414	737	129	20	99	43	1,720

Note: The sample covers only observations with a given or calculable standard error.

B.3 Distribution of Estimates by Year of Publication

Table 10: Year of Publication and Published Type

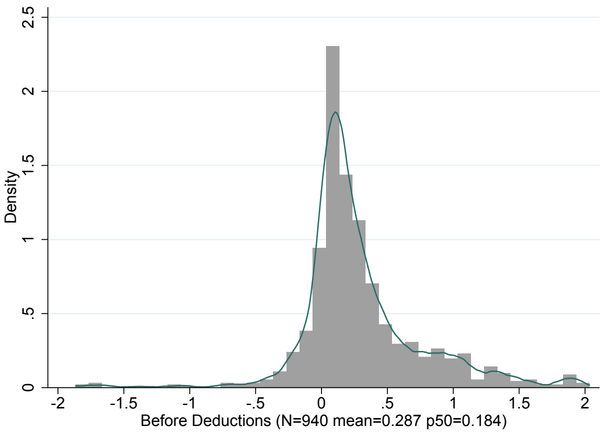
Year of Publication	**Working Paper**	**Published**	**Total**
1987	0	14	14
1998	12	0	12
1999	0	20	20
2000	0	39	39
2001	0	33	33
2002	0	35	35
2003	0	91	91
2004	11	0	11
2005	0	91	91
2006	15	0	15
2007	0	99	99
2008	72	10	82
2009	39	0	39
2010	21	151	172
2011	0	77	77
2012	27	32	59
2013	0	18	18
2014	25	191	216
2015	59	78	147
2016	50	82	124
2017	58	0	58
2018	8	36	44
2019	182	44	226
Total	579	1141	1,720

Note: The sample covers only observations with a given or calculable standard error.

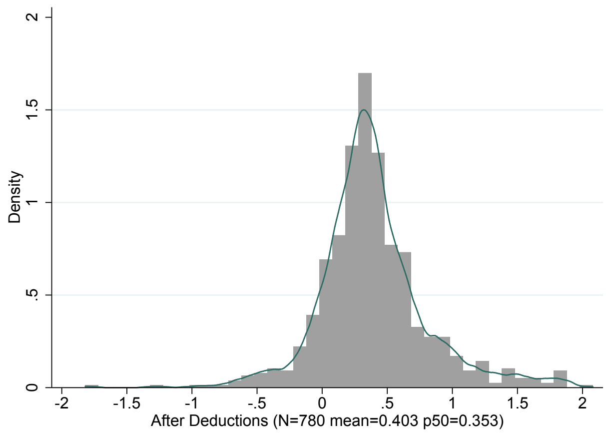
C Distribution of Elasticities and Details on Explanatory Variables

C.1 Distribution of Elasticities

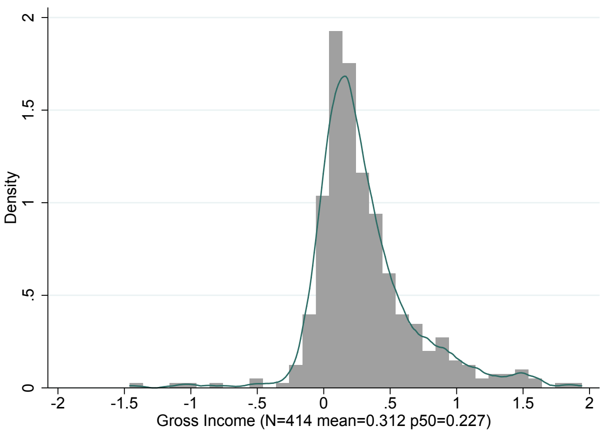
Figure 4: Distribution of Elasticities



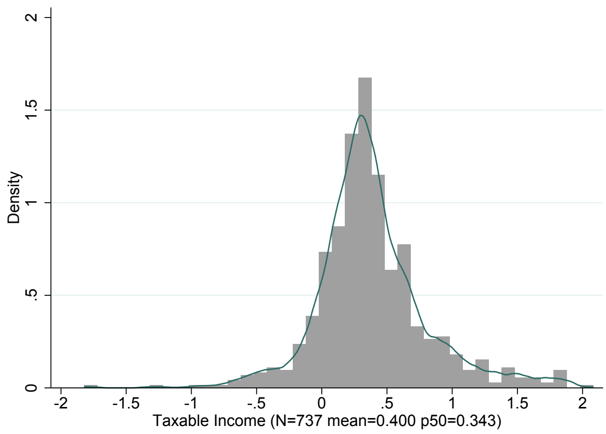
(a) Before Deductions



(b) After Deductions



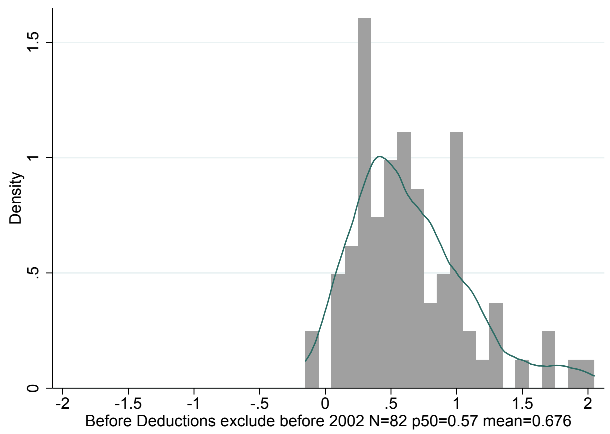
(c) Gross Income Elasticities



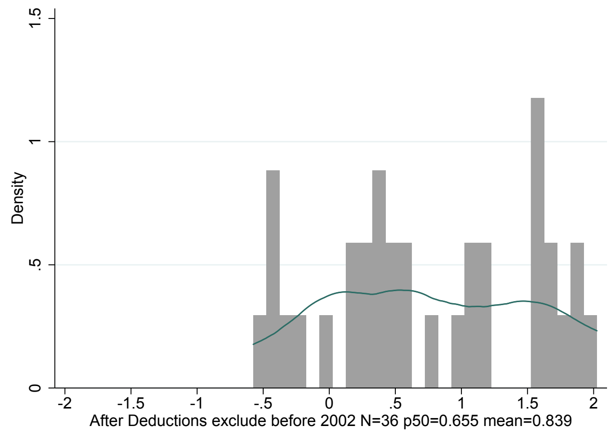
(d) Taxable Income Elasticities

Note: The data cover only observations with a given standard error or z-statistic. I restrict the sample to elasticity estimates that belong to the (a) before deductions subsample or (b) the after deduction subsample. Subfigures (c) and (d) are based on a narrower definition (gross or taxable income respectively).

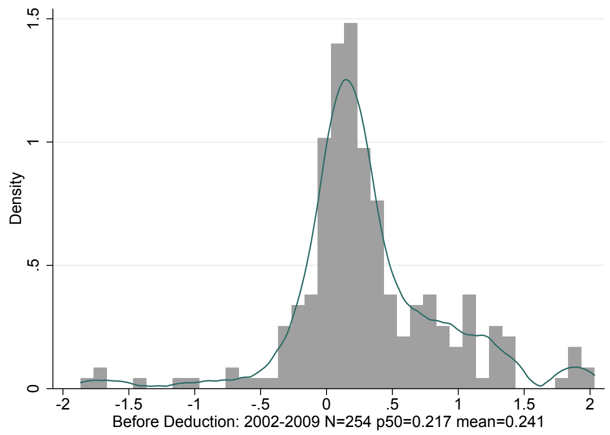
Figure 5: Distribution of Estimates by Publication Decade.



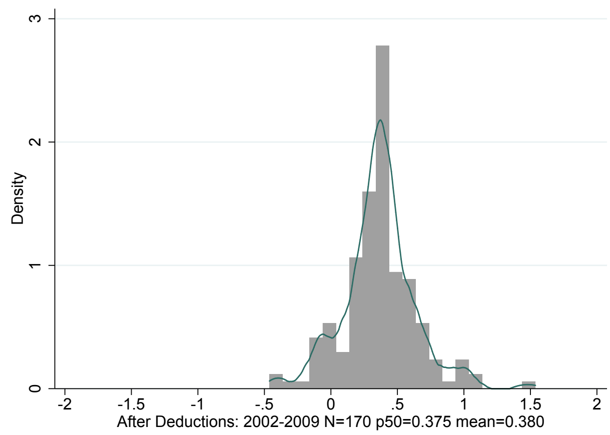
(a) Before Deductions <2002



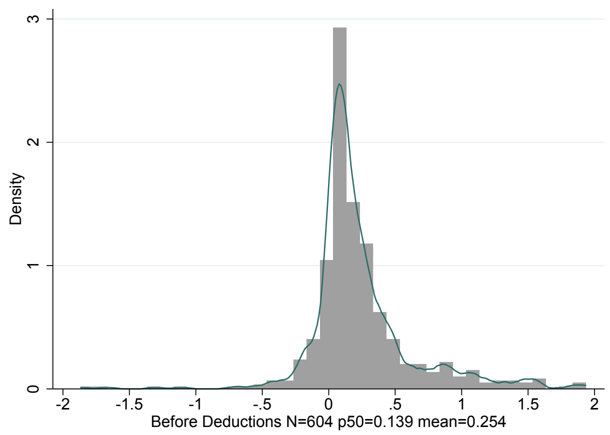
(b) After Deductions <2002



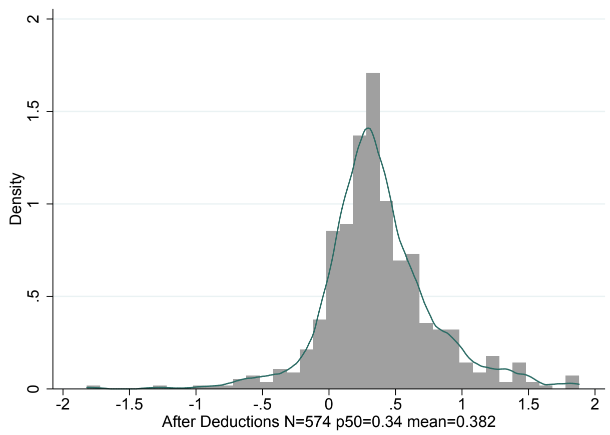
(c) Before Deductions >=2002 and <2009



(d) After Deductions >=2002 and <2009



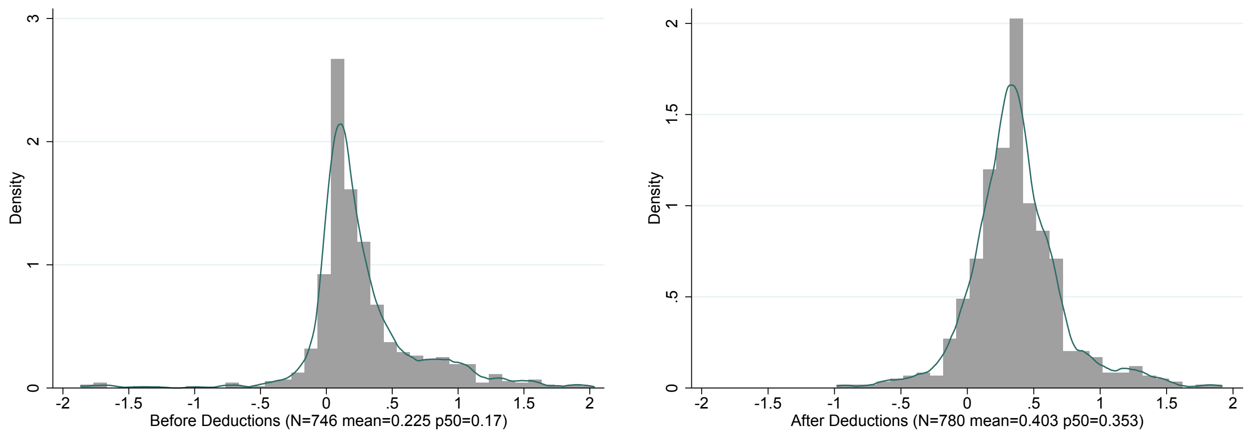
(e) Before Deductions >=2009



(f) After Deductions >=2009

Note: All graphs plot the distribution of elasticities by subsample and publication decade.

Figure 6: Distribution of Estimates (only income control(s)).



(a) Before Deductions (b) After Deductions

Note: Both graphs plot the distribution of elasticities that are derived with a specification using income control(s).

C.2 Explanatory Variables: Details

Regression technique: Most approaches use an Instrument for Δ NTR = $ln\left[\frac{(1-\tau_{it}(z_{it}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$ to achieve a causal relationship:

IV: mechanical tax rate changes: $\Delta ln(1-\tau_{it}^{p}) = ln\left[\frac{(1-\tau_{it}^{p}(z_{it-k}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$, where τ_{it}^{p} is the marginal tax rate that an individual would face given her synthetic income. Example: In year 3, τ_{it}^{p} would be calculated based on income of year two (assume time length of one year). Introduced by Auten and Carroll (1999) / Gruber and Saez (2002) and often referred to as the most standard specification.

IV: (lagged) mechanical tax rate changes: $\Delta ln(1-\tau_{it}^{p,lag})$, where $\tau_{it}^{p,lag}$ is based on income further in the past. $\Delta ln(1-\tau_{it}^{p}, lag) = ln\left[\frac{(1-\tau_{it}^{p}(z_{it-k-lag}))}{(1-\tau_{it-k}(z_{it-k}))}\right]$

IV: other: This category summarizes all other instruments. (1) Blomquist and Selin (2010): They use a single difference and an imputed taxable income $\hat{z_{it}}$ to calculate their instrument: $\left(\frac{1-\tau_{it}(\hat{z_{it}})}{1-\tau_{it-k}(\widehat{z_{it-k}})}\right)$. (2) Burns and Ziliak (2017): use a grouping estimator/instrument. (3) Carey et al. (2015): Two instruments based on a time period with no tax changes to estimate dynamics of taxable income. (4) Carroll (1998): proxy for permanent income and calculate synthetic tax rate. (5) Ericson et al. (2015): instrument based on individual/household-specific variables/no measure of previous or future taxable income. (6) Harju and Matikka (2016): use Gruber and Saez (2002) and Weber (2014) but include separate NTR for wage and dividend (plus, separate instruments). (7) Homlund and Söderström (2011): use a dynamic model to explicitly measure short and long run responses. (8) Looney and Singhal (2006): NTR change based on family income stays the same; predict the change in marginal tax rates faced by families assuming that family income remains constant in real terms between year 1 and year 2. (9) Matikka (2018): use changes in flat municipal income tax rates as an instrument for overall changes in marginal tax rates. This instrument is not a function of individual income, which is the basis for an exogenous instrument. (10) Gelber (2014) explicitly control for NTR for wife and husband and extend the most standard specification to allow each spouse's earnings to depend not only on his or her own tax rate and unearned income, but also on the tax rate and unearned income of the other spouse.

DID and IV: Combination of a classical DID and an IV- estimation procedure. The instrument is a binary dummy variable. It determines treatment and control. (e.g. Saez, 2003 or Kopczuk, 2015)

DID classic.

Income Controls: For the majority of coded specifications, there is no information available about what type of income (e.g. gross or taxable) is used.
Auten and Carroll (1999): '**Auten Carroll**' describes the use of log base year income $ln(z_{i,t-k})$ as an income control.
Mostly old studies and robustness checks deliver estimates that use no income control (**none)** at all.
Gruber and Saez (2002): '**Gruber Saez**' defines the inclusion of a spline of base year income as an income control.
Kopczuk (2005): '**Kopczuk**' defines the inclusion of two income control variables. The deviation of log base year income and lagged base year income and lagged base year income separately. To be more precise: $ln(z_{i,t-k-1})$, $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, spline of $ln(z_{i,t-k-1})$, spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of $ln(z_{i,t-k-1})$ and $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of $ln(z_{i,t-k})$ and spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$, combination of spline of $ln(z_{i,t-k})$ and $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$ and combination of spline of $ln(z_{i,t-k})$ and spline of $ln(z_{i,t-k}) - ln(z_{i,t-k-1})$.
The category '**other**' involves all other kinds of income controls. Example: Burns and Ziliak (2017) use a cohort-state-year income control in some specifications.

Difference Length The term difference length defines the time window k. If researchers relate 2005 to 2002, the time window will be 3 years.

Weighting by Income: This is a dummy variable that indicates whether (primary) estimation results are weighted by income.

Sample Restrictions:
Age Cutoff: It is a dummy variable that indicates whether an age cutoff is used.
Income Cutoff: I create subcategories: 0-10k, 10-12k, 12-31k and none. Some researchers do not apply any kind of income restrictions. However, sometimes it is not clear if they simply do not mention them, applied no income restriction on purpose or if their dataset considers a subgroup of tax-units in the first place. It often remains unclear what type of income is used (e.g. taxable or gross) to restrict the sample. I coded the values in national currency and recalculated them in US-Dollar. Purchasing power parities do not lead to different results.
Employment type: I distinguish between no restriction with respect to employment type (none), only wage earner, and only self employed individuals.
Marital Status: I distinguish between no restriction with respect to marital status (none), only married tax-units and only singles.

Variations across time and country:
Country Group: USA, Scandinavia (Denmark, Norway, Sweden) and Rest (Canada, Finland, France, Germany, Hungary, Netherlands, New Zealand, Poland, Spain)
Mean year in study data: I calculate the (rounded) mean year of observation based on time start and time end of dataset.
Estimation/Data Decade: I used the mean year of the study data and assigned the respective decade: $<$ 1990, 1990-2000 and $>=$ 2000.

Publication Characteristics:
Publication Decade: 2001-1010, $< =$ 2000 and $>$ 2011.
Published Type: I distinguish between (1) published in a peer reviewed journal and (2) Working Paper.

Extension: Contextual Variables: For a particular estimate, I compare start and end year of (restricted) data period and add the tax related characteristics. Economy related characteristics are merged via the mean year of observation.
Tax Reform Characteristics: It is difficult and almost impossible to code precisely if taxes are increased, and if so, by how much. As an example, think of an estimate that uses data from 2001 to 2010 and exploits three tax changes at different points in the income distribution which differ additionally in magnitude. Therefore, I decided to focus on: (1) introduction of a top tax bracket.
Intro of top tax bracket: information if reform involves an introduction of top. Source: Paper itself plus OECD Tax Database
Economy related characteristics merged via link to mean year of observation (= use start and end year of (restricted) data period for collected primary estimate:
Gini (disposable income, post taxes and transfers)/Income Definition till 2011. To improve (regression) interpretation, I standardized the Gini Coefficient by multiplying it with 100. Remark: These tables are updated on a regular basis. No data is available for China and South Africa. Source: http://stats.oecd.org (07.11.2016/18.06.2019)
Top Income Shares: Pre-tax national income share held by a given percentile group (here top 1% and top 10%). Pre-tax national income is the sum of all pre-tax personal income flows accruing to the owners of the production factors, labour and capital, before taking into account the operation of the tax/transfer system, but after taking into account the operation of pension system. No data available for Israel and South Africa. Source: World Inequality Database (extracted 16.07.2018/18.06.2019)
Unemployment Rate: The unemployment rate is the number of unemployed people as a percentage of the labour force, with the latter consisting of the unemployed plus those in paid or self-employment. Unemployed people are those who report that they are out of work, that they are available for work and that they have taken active steps to find work in the last four weeks. When unemployment is high, some people become discouraged and stop looking for work; they are then excluded from the labour force. This implies that the unemployment rate may fall, or stop rising, even though there has been no underlying improvement in the labour market. For South Africa and China no data available. (Source: OECD, Short-Term Labour Market Statistics; extracted 17.07.2018/18.06.2019.)
Fraction self-employed: fraction self-employed is defined crudely as all non employees (self-employed, employers, and non classifiable workers) as a fraction of the workforce. For Israel no data available. Source: Kleven - How Can Scandinavians Tax So Much? (2014, Journal of Economic Perspectives)
Modern taxes/GDP: Kleven et al. (2016) decompose the tax take (=tax/GDP) into modern and traditional taxes. Modern taxes include individual and corporate income taxes, payroll taxes and social security contributions, and value added taxes. Traditional taxes include all the other taxes. For Israel no data available. Source: Kleven et al. - Why Can Modern Governments Tax So Much? (2016, Economica)

Table 11: Descriptive Statistics: Categories of Heterogeneity

	Before Deductions (BD) (N=940) # studies	**After Deductions (AD) (N=780)** # studies
Estimation Techniques		
Regression technique		
IV: mechanical tax rate changes	32	32
IV: (lagged) mechanical tax rate changes	9	12
IV: other	6	11
DID and IV	7	3
classic DID	1	4
Income Control		
Auten Carroll (1999)	23	23
none	28	28
Gruber Saez (2002) spline	18	14
Kopczuk (2005) type	19	21
other	7	4
Difference Length		
3 years	20	24
1 year	25	24
2 years	13	14
4+ years	9	8
Weighted by Income	16	15
Sample Restrictions		
Age Cutoff	23	27
Income Cutoff		
0-10k	15	17
none	11	11
10k-12k	17	11
12-31k	19	15
$>$ 31k	23	21
Variations across Countries and Time		
Country Group		
USA	20	19
Scandinavia	5	67
other countries	16	20
Mean year in study data		
Estimation decade		
$<$ 1999	15	17
1990 - 2000	15	10
$>$ 2000	16	23
Publication Characteristics		
Publication decade		
2001-2010	11	15
$<=$ 2000	3	3
$>$ 2011	27	28
Published Type		
published in peer reviewed journal	25	27
working paper	16	19

Note: see text for description of sample. I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic. Reference categories are given in italics.

D Additional Sample Restrictions - Before Deductions (BD) and After Deductions

Researchers often conduct subgroup analysis by *marital status* or *employment type*. Single taxpayers might respond differently than married couples and it is obvious that a self-employed person has more control over his or her income compared to someone receiving only wage income.

Table 12: Descriptive Statistics: Sample Restrictions

	Before Deductions (BD) (N=940)		**After Deductions (AD) (N=780)**	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
Sample Restrictions				
Employment type				
none	0.710	0.454	0.877	0.329
wage earner	0.218	0.413	0.040	0.195
self-employed	0.072	0.259	0.083	0.277
Marital Status				
none	0.845	0.362	0.858	0.350
married	0.110	0.313	0.092	0.290
single	0.046	0.209	0.050	0.218

Note: see text for description of sample. I present descriptive results separately for two subsamples: before (BD) and after deductions (AD). The sample covers only observations with a given standard error or t-statistic.

In line with expectations, a BD elasticity estimated on a subsample of only wage earners leads to a lower elasticity compared to a specification with no restriction on employment type. Greater coverage of third party information reporting and the associated lower evasion opportunities might be a reason (Kleven et al., 2011). If primary studies restrict their sample according to marital status, it appears that single taxpayers reveal a lower BD elasticity compared to no restriction.

Table 13: WLS before deductions results with add. sample restrictions

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.060*	0.054*	0.061*	0.055*	0.028**	0.025*
	(0.031)	(0.029)	(0.032)	(0.030)	(0.013)	(0.015)
IV-other	0.075	0.081*	0.070	0.078*	0.074	0.107*
	(0.056)	(0.044)	(0.055)	(0.042)	(0.053)	(0.056)
DID-IV	0.298***	0.224**	0.291***	0.218*	0.319***	0.313***
	(0.053)	(0.105)	(0.058)	(0.109)	(0.046)	(0.075)
DID-classic	0.332***	0.068	0.309***	0.049	0.184***	0.149**
	(0.059)	(0.132)	(0.078)	(0.137)	(0.060)	(0.065)
Income Control (omitted: Auten Carroll)						
none	-0.213***	-0.212***	-0.213***	-0.211***	-0.207***	-0.207***
	(0.024)	(0.025)	(0.024)	(0.025)	(0.029)	(0.029)
Gruber Saez Spline	-0.020***	-0.021***	-0.021***	-0.022***	-0.014**	-0.016***
	(0.005)	(0.007)	(0.006)	(0.007)	(0.006)	(0.005)
Kopczuk	-0.017**	-0.014**	-0.018**	-0.015**	-0.012*	-0.010**
	(0.007)	(0.005)	(0.008)	(0.006)	(0.006)	(0.005)
other	-0.034**	-0.029**	-0.033*	-0.029*	-0.012	-0.009
	(0.017)	(0.013)	(0.018)	(0.015)	(0.010)	(0.009)
Difference Length (omitted: 3-years)						
1 year	0.060	0.033	0.058	0.032	0.031	0.012
	(0.063)	(0.045)	(0.062)	(0.044)	(0.051)	(0.040)
2 years	-0.013	-0.030*	-0.015	-0.033*	-0.042***	-0.035***
	(0.021)	(0.016)	(0.021)	(0.019)	(0.010)	(0.008)
4 years and more	0.082*	0.068**	0.084*	0.068**	0.012	0.026
	(0.042)	(0.030)	(0.043)	(0.029)	(0.020)	(0.021)
Sample Restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied		-0.282**		-0.278**		-0.267
		(0.122)		(0.123)		(0.174)
Income Cutoff applied (omitted: 0-10k)						
none		0.018		0.019		-0.020*
		(0.021)		(0.022)		(0.010)
10k-12k		0.024		0.026		-0.015**
		(0.016)		(0.016)		(0.007)
12k-31k		0.009		0.009		0.007
		(0.007)		(0.010)		(0.008)
>31k		0.021		0.023		-0.005
		(0.017)		(0.019)		(0.012)
Employment Type (omitted: no restriction)						
wage self=0			0.000	0.000		
wage earner			-0.008*	-0.005		
			(0.005)	(0.005)		
self-employed			0.006	0.007		
			(0.009)	(0.011)		
Marital Status (omitted: no restriction)						
married			0.021	0.022		
			(0.031)	(0.036)		
sinlge			0.012	0.009		
			(0.030)	(0.028)		
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia					0.074	0.239*
					(0.081)	(0.123)
other countries					0.191**	0.343***
					(0.081)	(0.126)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					0.226	0.426**
					(0.141)	(0.207)
after 2010					-0.254**	-0.205***
					(0.098)	(0.073)
Constant	0.073***	0.351***	0.079***	0.350***	0.244***	0.296***
	(0.007)	(0.123)	(0.009)	(0.123)	(0.043)	(0.054)
Observations	940	940	940	940	940	940
Adjusted R^2	0.566	0.615	0.566	0.615	0.628	0.655

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: WLS after deductions results with add. sample restrictions

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)
Estimation Technique:						
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.409***	0.420***	0.461***	0.455***	0.232***	0.207***
	(0.088)	(0.061)	(0.047)	(0.054)	(0.048)	(0.074)
IV-other	-0.265*	-0.246**	-0.224***	-0.227*	0.403*	0.197
	(0.145)	(0.118)	(0.078)	(0.130)	(0.230)	(0.218)
DID-IV	-0.590**	-0.702**	-0.423***	-0.583**	-0.152	-0.289
	(0.224)	(0.281)	(0.155)	(0.240)	(0.403)	(0.475)
DID-classic	-0.188	-0.189	-0.152	-0.162	-0.167	-0.178
	(0.372)	(0.363)	(0.320)	(0.324)	(0.323)	(0.305)
Income Control (omitted: Auten Carroll)						
none	0.108	0.045	0.100	0.043	-0.225	-0.249
	(0.078)	(0.089)	(0.069)	(0.096)	(0.176)	(0.159)
Gruber Saez Spline	-0.100	-0.137**	-0.086	-0.120*	-0.087	-0.119
	(0.068)	(0.068)	(0.054)	(0.067)	(0.066)	(0.088)
Kopczuk-type	-0.371***	-0.387***	-0.375***	-0.383***	0.027	0.025
	(0.043)	(0.075)	(0.047)	(0.087)	(0.068)	(0.104)
other	-0.195**	-0.331**	-0.174**	-0.308**	0.108	0.048
	(0.075)	(0.132)	(0.085)	(0.136)	(0.074)	(0.124)
Difference Length (omitted: 3-years)						
1 year	-0.048	0.073	-0.049	0.066	-0.001	0.119
	(0.106)	(0.074)	(0.121)	(0.085)	(0.127)	(0.090)
2 years	0.033	0.019	0.021	0.008	0.043	0.057
	(0.086)	(0.119)	(0.091)	(0.117)	(0.102)	(0.105)
4 years and more	0.285	0.182	0.290	0.188	-0.329	-0.362
	(0.191)	(0.212)	(0.189)	(0.210)	(0.247)	(0.242)
Sample restrictions:						
Age Cutoff applied (omitted: no restriction)						
Age Cutoff applied		0.252**		0.245**		0.140
		(0.113)		(0.113)		(0.124)
Income Cutoff applied (omitted: 0-10k)						
none		0.154***		0.147**		0.254***
		(0.054)		(0.057)		(0.087)
10k-12k		0.109		0.099		0.353
		(0.090)		(0.088)		(0.236)
12k-31k		0.111*		0.105		0.068
		(0.063)		(0.063)		(0.059)
>31k		0.468		0.462		0.518
		(0.424)		(0.423)		(0.353)
Employment Type (omitted: no restriction)						
wage earner			-0.007	-0.019		
			(0.050)	(0.031)		
self-employed			-0.274***	-0.208*		
			(0.052)	(0.105)		
Marital Status (omitted: no restriction)						
married			-0.074	-0.035		
			(0.096)	(0.071)		
single			0.010	0.012		
			(0.098)	(0.090)		
Variation across countries and time:						
Country Group (omitted: USA)						
Scandinavia					0.121	0.410
					(0.112)	(0.305)
other countries					0.416***	0.632**
					(0.136)	(0.304)
(Publication) Decade (omitted: 2001-2010)						
prior to 2001					1.060*	1.164*
					(0.599)	(0.662)
after 2010					-0.468***	-0.500***
					(0.161)	(0.173)
Constant	0.445***	0.208***	0.455***	0.222***	0.376***	-0.019
	(0.040)	(0.066)	(0.041)	(0.068)	(0.098)	(0.272)
Observations	780	780	780	780	780	780
Adjusted R^2	0.405	0.479	0.414	0.483	0.553	0.621

Note: Columns (1) to (6) estimated using WLS with the inverse of an estimate's variance as analytical weights. Reported coefficients need to be interpret as a deviation from the reference category (in bold). Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

E Contextual Factors - Full Results

E.1 Contextual Factors - Before Deductions (BD) - Full Results

Table 15: WLS before deductions - Contextual Variables

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reg. Technique (omitted: IV: Δ mechanical tax rate)							
IV: lagged Δ mechanical tax rate	0.063	0.040***	0.048**	0.058**	0.066*	0.044**	0.038**
	(0.039)	(0.014)	(0.019)	(0.028)	(0.036)	(0.019)	(0.017)
IV-other	0.095**	0.090	0.063	0.073	0.181*	0.050	0.077
	(0.037)	(0.057)	(0.059)	(0.058)	(0.106)	(0.057)	(0.061)
DID-IV	0.297***	0.272***	0.271***	0.291***	0.318***	0.296***	0.253***
	(0.055)	(0.068)	(0.077)	(0.060)	(0.090)	(0.050)	(0.083)
DID-classic	0.325***	0.270***	0.267***	0.319***	0.337***	0.359***	0.134*
	(0.073)	(0.066)	(0.077)	(0.066)	(0.076)	(0.048)	(0.075)
Income Control (omitted: Auten Carroll)							
none	-0.213***	-0.213***	-0.214***	-0.214***	-0.206***	-0.211***	-0.211***
	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)	(0.026)	(0.025)
Gruber Saez Spline	-0.020***	-0.011	-0.017***	-0.019***	-0.018**	-0.022***	-0.013*
	(0.005)	(0.009)	(0.005)	(0.005)	(0.008)	(0.006)	(0.008)
Kopczuk-type	-0.016**	-0.013*	-0.016**	-0.016**	-0.018**	-0.021**	-0.011*
	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.006)
other	-0.035*	-0.065***	-0.053**	-0.043*	-0.044***	0.002	-0.047***
	(0.017)	(0.020)	(0.022)	(0.024)	(0.014)	(0.010)	(0.013)
Difference Length (omitted: 3-years)							
1 year	0.066	0.057	0.063	0.064	0.058	0.035	0.049
	(0.079)	(0.060)	(0.066)	(0.066)	(0.079)	(0.049)	(0.056)
2 years	-0.013	-0.005	-0.004	-0.008	-0.012	-0.040*	-0.014***
	(0.021)	(0.007)	(0.020)	(0.020)	(0.016)	(0.024)	(0.004)
4 years and more	0.083*	0.047**	0.083*	0.086*	0.068*	0.063**	0.038*
	(0.043)	(0.020)	(0.043)	(0.044)	(0.035)	(0.030)	(0.021)
Additional Variables							
Intro top bracket	-0.027						
	(0.078)						
Gini Coefficient		0.008***					
		(0.002)					
Top 10%			0.814*				
			(0.442)				
Top 1%				0.330			
				(0.448)			
Unemployment Rate					-0.007		
					(0.004)		
Fraction of self-employed						0.016***	
						(0.006)	
modern taxes (in 2005)							-0.010***
							(0.002)
Constant	0.073***	-0.118***	-0.142	0.047	0.113***	-0.081	0.460***
	(0.006)	(0.043)	(0.114)	(0.032)	(0.022)	(0.054)	(0.084)
Observations	940	931	912	912	854	915	921
Adjusted R^2	0.566	0.614	0.585	0.576	0.569	0.611	0.614

Columns (1) to (7) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For observations that are based on classic DID approach, I do not have information of the share of self employed that correspond to the respective mean year of observation.

E.2 Contextual Factors - After Deductions (AD) - Full Results

Table 16: WLS after deductions - Contextual Factors

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Reg. Technique (omitted: IV: Δ mechanical tax rate)							
IV: (lagged) Δ mechanical tax rate	0.410***	0.422***	0.287***	0.304***	0.393***	0.472***	0.510***
	(0.088)	(0.091)	(0.094)	(0.108)	(0.058)	(0.116)	(0.111)
IV-other	-0.265*	-0.279	-0.087	-0.016	-0.038	-0.300*	-0.391**
	(0.145)	(0.179)	(0.147)	(0.161)	(0.135)	(0.158)	(0.175)
DID-IV	-0.591**	-0.596**	-0.618***	-0.699***	-0.042	-0.653***	-0.498*
	(0.223)	(0.225)	(0.084)	(0.109)	(0.178)	(0.222)	(0.296)
DID-classic	-0.189	-0.201	-0.011	-0.009	-1.130**	-0.264	-0.305
	(0.372)	(0.398)	(0.363)	(0.340)	(0.482)	(0.377)	(0.376)
Income Control (omitted: Auten Carroll)							
none	0.107	0.107	0.029	-0.013	0.029	0.021	0.045
	(0.078)	(0.077)	(0.083)	(0.097)	(0.108)	(0.116)	(0.096)
Gruber Saez Spline	-0.100	-0.080	-0.112*	-0.107*	-0.043	-0.102	-0.084
	(0.068)	(0.064)	(0.064)	(0.060)	(0.049)	(0.085)	(0.067)
Kopczuk-type	-0.371***	-0.385***	-0.290***	-0.352***	-0.333***	-0.454***	-0.493***
	(0.043)	(0.111)	(0.061)	(0.053)	(0.067)	(0.122)	(0.123)
other	-0.190*	-0.240	-0.087	-0.147	-0.304*	-0.374*	-0.368*
	(0.096)	(0.151)	(0.099)	(0.122)	(0.158)	(0.189)	(0.184)
Difference Length (omitted: 3-years)							
1 year	-0.048	-0.042	-0.074	-0.094	-0.066	0.013	-0.009
	(0.106)	(0.114)	(0.105)	(0.100)	(0.122)	(0.100)	(0.099)
2 years	0.035	0.061	-0.060	-0.063	-0.017	0.088	0.100
	(0.084)	(0.133)	(0.081)	(0.090)	(0.095)	(0.121)	(0.126)
4 years and more	0.288	0.267	0.041	-0.042	0.430	0.209	0.482*
	(0.188)	(0.211)	(0.245)	(0.266)	(0.436)	(0.200)	(0.253)
Additional Variables							
Intro top bracket	-0.016						
	(0.132)						
Gini Coefficient		-0.002					
		(0.014)					
Top 10%			3.563**				
			(1.536)				
Top 1%				7.709**			
				(3.202)			
Unemployment Rate					0.067*		
					(0.039)		
Fraction of self-employed						-0.022	
						(0.023)	
modern taxes (in 2005)							0.016
							(0.012)
Constant	0.450***	0.513	-0.572	-0.159	-0.088	0.746**	-0.060
	(0.116)	(0.424)	(0.435)	(0.243)	(0.315)	(0.349)	(0.363)
Observations	780	767	771	771	703	771	780
Adjusted R^2	0.404	0.410	0.455	0.469	0.468	0.425	0.426

Columns (1) to (7) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

For observations that are based on classic DID approach, I do not have information of the share of self employed that correspond to the respective mean year of observation.

F Sensitivity Analysis and Robustness Checks

F.1 Sensitivity Analysis

In this section, I limit the number of estimates along various dimensions: (i) I drop studies that are released prior to 2002, (ii) I consider only published articles or (iii) only US studies and (iv) I only consider taxable income elasticities. Results are presented in Table 17 and they vary slightly compared to the baseline results when I consider only published articles and only US studies. For US studies, the constant for BD elasticities is larger and smaller for AD elasticities compared to the baseline results shown in Table 2 and 3 (column 2).[22] Moreover, the degree of influence of other factors changes. The use of (lagged) mechanical tax rate changes lead to an increase of 0.541 compared to an approach that relies on mechanical tax rate changes as an instrument. On the other hand DID and DID IV does not make a big difference compared to an approach using the standard mechanical tax rate changes instrument. The coefficient of DID-classic is very large but mainly driven by older studies (reported < 2002).

Table 17: Sensitivity Analysis: Different Sample Restrictions

Dependent Variable:	drop studies prior to 2002		(only) Published		(only) US studies		(only) Taxable
Income Elasticity ...	BD	AD	BD	AD	BD	AD	Income
Reg. Technique (omitted: IV:Δ mech. tax rate))							
IV: (lagged) Δ mech. tax rate	0.060*	0.410***	0.060*	0.420***	0.395**	0.271**	0.409***
	(0.031)	(0.088)	(0.031)	(0.086)	(0.155)	(0.123)	(0.088)
IV-other	0.055	-0.261*	0.055	0.690***	-0.003	0.309***	-0.274*
	(0.054)	(0.147)	(0.053)	(0.117)	(0.094)	(0.093)	(0.142)
DID-IV	0.295***	-0.248**	0.290***	0.239	-0.026	0.115*	-0.751***
	(0.055)	(0.113)	(0.055)	(0.239)	(0.120)	(0.064)	(0.220)
DID-classic	0.332***	-0.225	0.337***	0.076	-0.054	1.302***	-1.432***
	(0.059)	(0.395)	(0.058)	(0.173)	(0.116)	(0.128)	(0.474)
Income Control (omitted: Auten Carroll)							
none	-0.212***	0.103	-0.215***	-0.873***	-0.083	0.012	0.124
	(0.025)	(0.079)	(0.024)	(0.127)	(0.140)	(0.171)	(0.080)
Gruber Saez Spline	-0.019***	-0.105	-0.020***	-0.070	-0.150*	0.020	-0.089
	(0.005)	(0.069)	(0.005)	(0.054)	(0.074)	(0.073)	(0.065)
Kopczuk-type	-0.015**	-0.377***	-0.017**	-0.300***	-0.240**	-0.062	-0.360***
	(0.006)	(0.045)	(0.007)	(0.043)	(0.087)	(0.067)	(0.050)
other	-0.033*	-0.195**	-0.033*	-0.254***	-0.134	0.017	-0.186**
	(0.017)	(0.076)	(0.017)	(0.042)	(0.082)	(0.123)	(0.075)
Difference Length (omitted: 3-years)							
1 year	0.060	-0.052	0.056	0.024	0.078	-0.155**	-0.046
	(0.064)	(0.109)	(0.063)	(0.062)	(0.102)	(0.073)	(0.107)
2 years	-0.013	0.025	-0.015	0.081	-0.137	-0.079	0.032
	(0.021)	(0.086)	(0.021)	(0.104)	(0.161)	(0.061)	(0.087)
4 years and more	0.081*	0.125	0.081*	0.089	0.147*	0.023	0.431
	(0.043)	(0.175)	(0.043)	(0.335)	(0.083)	(0.137)	(0.285)
Constant	0.072***	0.451***	0.073***	0.369***	0.310***	0.258**	0.434***
	(0.006)	(0.042)	(0.007)	(0.045)	(0.066)	(0.093)	(0.048)
Observations	858	744	822	701	464	415	737
Adjusted R^2	0.571	0.407	0.592	0.623	0.063	0.363	0.434

Note: BD refers to the before deductions subsample and AD to the after deductions subsample. All results are based on Weighted Least Squares (WLS) with the inverse of an estimate's variance as analytical weights. The baseline specification involves only controls for estimation technique (regression technique, income control and difference length). Standard errors (in parentheses) are clustered at the study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[22] The results within the US subsample but also the baseline results remain remarkably robust even when I exclude all estimates extracted from Weber(2014).

F.2 Robustness Checks: Different Estimation Techniques

The upper (lower) part of the Table displays results based on the BD (AD) subsample. Column (1) display the baseline results obtained in column (2) of Tables 2 and 3. In Column (2), I present results based on a random effects meta-regression technique. The weights in the baseline WLS represent only the within study variance and neglect any possible between study variance. In contrast the estimation used here, it is equivalent to the baseline WLS with an additive between study component in the denominator of the weights. **?** show that WLS is superior to conventional random-effects meta-regression estimation. In case of publication bias, in particular, WLS always reveals a smaller bias than the random effects model. Moreover, random effects estimates are highly sensitive to the accuracy of the estimate of the between study variance.

For illustration, results based on a simple OLS are presented in column (4). Since we observe large heteroscedasticity among estimates, an OLS procedure is never appropriate in a meta analysis. To increase efficiency, a WLS procedure is always preferable.

Column (5) shows results that are based on WLS with weights that are based on the inverse of the share of observations per study in relation to the full sample. Given that my collected sample does not consist only of one estimate per study but of all available estimates a particular study provides, there's a risk that the baseline results are driven only by a small number of studies that offer a lot of estimates.

It seems reasonable to assume that extracted estimates themselves are influenced by their sample size. For instance, a dataset that almost covers the entire population might produce a different estimate and standard error compared to a dataset of a few hundred observations. In column (6) I weight each primary estimate with the sample size of the respective study. The difference between those results compared to a standard WLS with precision as a weight should be small, since the sampling error is to large extent determined by the respective sample size.

The BD subsample is based on 38 studies and the AD subsample on 37 studies. To check whether clustering in the meta-analysis produces misleading inferences, I apply a wild-cluster bootstrap procedure proposed by Cameron et al. (2008) for improved inference with only few cluster (see Column (3)).

Table 18: Robustness Checks: Different Estimation Techniques

Dependent Variable: Income Elasticity BEFORE deductions	(1) WLS	(2) META	(3) WILD	(4) OLS	(5) EQUAL	(6) NOBS
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.060*	0.104*	0.060	0.254	0.400	0.124***
	(0.031)	(0.059)	(0.061)	(0.264)	(0.335)	(0.038)
IV-other	0.075	-0.096*	0.075	-0.228*	-0.428***	-0.016
	(0.056)	(0.057)	(0.065)	(0.135)	(0.154)	(0.093)
DID-IV	0.298***	0.080*	0.298***	-0.289	-0.230	0.475***
	(0.053)	(0.048)	(0.000)	(0.247)	(0.166)	(0.107)
DID-classic	0.332***	-0.065	0.332***	-0.583	-0.501***	0.173*
	(0.059)	(0.300)	(0.000)	(0.385)	(0.182)	(0.101)
Income Control (omitted: Auten Carroll)						
none	-0.213***	-0.156***	-0.213***	0.276	-0.044	-0.183***
	(0.024)	(0.036)	(0.069)	(0.322)	(0.170)	(0.062)
Gruber Saez Spline	-0.020***	-0.152***	-0.020***	-0.325**	-0.190	-0.040*
	(0.005)	(0.034)	(0.007)	(0.127)	(0.213)	(0.024)
Kopczuk-type	-0.017**	-0.195***	-0.017***	-0.243*	-0.371**	-0.015
	(0.007)	(0.031)	(0.005)	(0.125)	(0.164)	(0.013)
other	-0.034*	-0.248***	-0.034	-0.266**	-0.413***	-0.114***
	(0.017)	(0.040)	(0.031)	(0.109)	(0.118)	(0.037)
Difference Length (omitted: 3-years)						
1 year	0.060	0.179***	0.060	0.158	0.281**	0.174
	(0.063)	(0.029)	(0.089)	(0.140)	(0.138)	(0.104)
2 years	-0.013	-0.059	-0.013	-0.121	-0.141	0.047
	(0.021)	(0.038)	(0.022)	0.113)	(0.156)	(0.032)
4 years and more	0.082*	0.014	0.082	-0.016	0.047	0.117***
	(0.042)	(0.033)	(0.125)	(0.138)	(0.136)	(0.035)
Constant	0.073***	0.292***	0.073***	0.404***	0.519***	0.078***
	(0.007)	(0.026)	(0.000)	0.128)	(0.136)	(0.017)
Observations	940	940	940	940	940	869
Adjusted R^2	0.566		0.566	0.020	0.065	0.114
Income Elasticity AFTER deductions	(1) WLS	(2) META	(3) WILD	(4) OLS	(5) EQUAL	(6) NOBS
Reg. Technique (omitted: IV: mechanical tax rate changes)						
IV: (lagged) mechanical tax rate changes	0.409***	0.294***	0.409***	0.326***	0.320**	0.445***
	(0.088)	(0.050)	(0.000)	(0.095)	(0.123)	(0.052)
IV-other	-0.265*	0.083	-0.265	0.181	0.411*	-0.108
	(0.145)	(0.052)	(0.293)	(0.123)	(0.226)	(0.127)
DID-IV	-0.590**	-0.129	-0.590	-0.104	-0.153	-0.146
	(0.224)	(0.081)	(0.530)	(0.125)	(0.160)	(0.093)
DID-classic	-0.188	0.578***	-0.188	0.551	0.814**	-0.144
	(0.372)	(0.071)	(0.278)	(0.331)	(0.401)	(0.296)
Income Control (omitted: Auten Carroll)						
none	0.108	-0.014	0.108*	-0.021	-0.276	0.030
	(0.078)	(0.044)	(0.059)	(0.130)	(0.206)	(0.065)
Gruber Saez Spline	-0.100	-0.000	-0.100	-0.056	-0.227	-0.126
	(0.068)	(0.045)	(0.080)	(0.068)	(0.169)	(0.100)
Kopczuk-type	-0.371***	-0.083**	-0.371***	-0.072	-0.193	-0.349***
	(0.043)	(0.041)	(0.120)	(0.088)	(0.177)	(0.094)
other	-0.195**	0.134	-0.195***	0.297	0.370	0.207
	(0.075)	(0.117)	(0.067)	(0.544)	(0.553)	(0.358)
Difference Length (omitted: 3-years)						
1 year	-0.048	0.018	-0.048	-0.044	0.088	-0.031
	(0.106)	(0.035)	(0.102)	(0.117)	(0.151)	(0.044)
2 years	0.033	0.091*	0.033	0.088	0.167*	-0.109***
	(0.086)	(0.053)	(0.074)	(0.105)	(0.087)	(0.039)
4 years and more	0.285	0.149**	0.285	0.264	0.651**	1.373**
	(0.191)	(0.066)	(0.229)	(0.221)	(0.251)	(0.644)
Constant	0.445***	0.295***	0.445***	0.317***	0.370***	0.484***
	(0.040)	(0.032)	(0.000)	(0.066)	(0.111)	(0.071)
Observations	780	780	780	780	780	728
Adjusted R^2	0.405		0.405	0.111	0.268	0.335

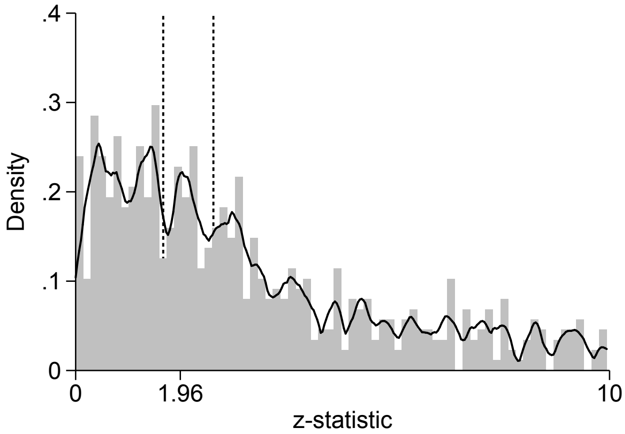
Except for column 3 standard errors (in parentheses) are clustered at the study level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The sample size in column (6) is lower because the sample size is not observed for every primary estimate.

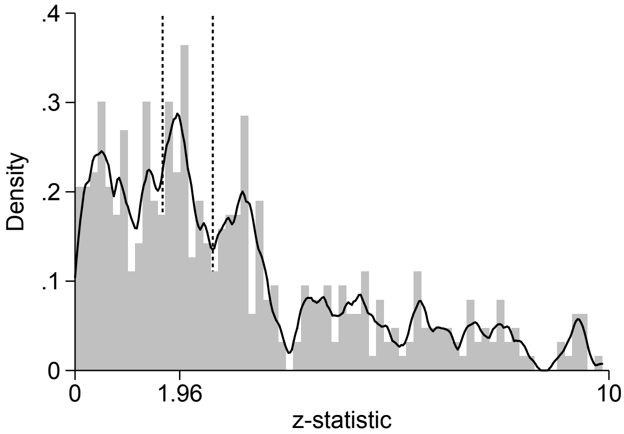
G Selective Reporting Bias: more information

G.1 Distribution of z-statistics - only with income controls

Figure 7: Distribution of z-statistics - only with income controls.



(a) Before Deductions



(b) After Deductions

Note: The left (right) figure is based on the before (after) deductions subsample. The 5% significance value (=1.96) is highlighted.

G.2 Selective Reporting Bias: BD - Full Results

Table 19: WLS before deductions: Publication Bias Full Results

Dependent Variable: Income Elasticity BEFORE deductions	(1)	(2)	(3)	(4)
Reg. Technique (omitted: IV: Δ mechanical tax rate)				
IV: lagged Δ mechanical tax rate	0.031*	0.029*	0.022	0.025
	(0.018)	(0.016)	(0.014)	(0.016)
IV-other	-0.165*	-0.164*	-0.235**	-0.196*
	(0.096)	(0.087)	(0.113)	(0.106)
DID-IV	0.198*	0.216**	0.197*	0.205**
	(0.101)	(0.095)	(0.103)	(0.098)
DID-classic	-1.052***	-0.797**	-0.199	-0.135
	(0.293)	(0.300)	(0.269)	(0.344)
Income Control (omitted: Auten Carroll)				
none	-0.211***	-0.210***	-0.211***	-0.210***
	(0.026)	(0.026)	(0.026)	(0.026)
Gruber Saez Spline	-0.020***	-0.018***	-0.020***	-0.019***
	(0.005)	(0.005)	(0.006)	(0.005)
Kopczuk-type	-0.018***	-0.016***	-0.019***	-0.017***
	(0.007)	(0.006)	(0.007)	(0.006)
other	-0.026*	-0.022**	-0.022*	-0.023*
	(0.013)	(0.010)	(0.012)	(0.013)
Difference Length (omitted: 3-years)				
1 year	0.034	0.030	0.024	0.029
	(0.052)	(0.051)	(0.046)	(0.049)
2 years	-0.026	0.005	-0.033**	-0.028*
	(0.016)	(0.011)	(0.014)	(0.016)
4 years and more	0.053*	0.046*	0.041	0.050
	(0.031)	(0.027)	(0.026)	(0.030)
Standard Error	3.654***	4.084***	0.972	0.652
	(0.719)	(0.845)	(0.812)	(0.988)
Journal impact factor		-0.012		
		(0.008)		
Std.Error* Impact Factor		-0.051		
		(0.035)		
Dummy if obs > median(obs)			0.771***	
			(0.279)	
Std.Error*D if obs > median(obs)			4.375***	
			(1.416)	
Dummy reported prior to 2009				0.575**
				(0.267)
Std.Error*D reported prior to 2009				3.726***
				(1.322)
Constant	0.876***	0.982***	0.477***	0.460**
	(0.159)	(0.186)	(0.138)	(0.181)
Observations	940	940	940	940
Adjusted R^2	0.614	0.624	0.628	0.627

Columns (1) to (4) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Included standard errors as explanatory variables are normalized. It allows an interpretation as standard deviation.

G.3 Selective Reporting Bias: AD - Full Results

Table 20: WLS after deductions Publication Bias Full Results

Dependent Variable: Income Elasticity AFTER deductions	(1)	(2)	(3)	(4)
Reg. Technique (omitted: IV: Δ mechanical tax rate)				
IV: lagged Δ mechanical tax rate	0.413***	0.205*	0.423***	0.426***
	(0.088)	(0.110)	(0.088)	(0.088)
IV-other	-0.264*	-0.066	-0.269*	-0.271*
	(0.143)	(0.167)	(0.140)	(0.138)
DID-IV	-0.577**	-0.390	-0.626**	-0.633**
	(0.230)	(0.246)	(0.258)	(0.299)
DID-classic	-0.186	-0.044	-0.266	-0.351
	(0.375)	(0.373)	(0.421)	(0.444)
Income Control (omitted: Auten Carroll)				
none	0.107	-0.020	0.125	0.134
	(0.075)	(0.097)	(0.086)	(0.084)
Gruber Saez Spline	-0.099	-0.139*	-0.069	-0.060
	(0.068)	(0.078)	(0.062)	(0.062)
Kopczuk-type	-0.372***	-0.052	-0.343***	-0.328***
	(0.042)	(0.092)	(0.055)	(0.054)
other	-0.193**	0.289	-0.168**	-0.160*
	(0.076)	(0.190)	(0.082)	(0.082)
Difference Length (omitted: 3-years)				
1 year	-0.048	-0.080	-0.030	-0.018
	(0.106)	(0.129)	(0.095)	(0.089)
2 years	0.034	0.031	0.046	0.061
	(0.089)	(0.114)	(0.090)	(0.093)
4 years and more	0.300	0.271	0.290*	0.354*
	(0.201)	(0.180)	(0.173)	(0.201)
Standard Error	-0.030	-0.834***	-0.223	-0.360
	(0.203)	(0.294)	(0.354)	(0.530)
Journal impact factor		0.030**		
		(0.014)		
Std.Error* Impact Factor		0.084***		
		(0.022)		
Dummy if obs > median(obs)			-0.066	
			(0.285)	
Std.Error*D if obs > median(obs)			0.113	
			(0.540)	
Dummy reported prior to 2009				-0.122
				(0.304)
Std.Error*D reported prior to 2009				0.217
				(0.614)
Constant	0.424**	-0.027	0.400**	0.416*
	(0.158)	(0.221)	(0.158)	(0.248)
Observations	780	780	780	780
Adjusted R^2	0.404	0.456	0.408	0.420

Columns (1) to (4) estimated using WLS. Standard errors (in parentheses) are clustered at the study level. Significance levels are * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Included standard errors as explanatory variables are normalized. It allows an interpretation as standard deviation.