Optimal Taxation and Enforcement with Market Power

Peter Choi

November 7, 2024

Frequently updated: Click here for latest version

Abstract

This paper explores the effect of market power on optimal tax and enforcement rates, finding that neither tax rates nor enforcement levels necessarily vary monotonically with market power. Higher effective taxation of an imperfectly competitive market compounds its distortions, but in some circumstances the associated welfare cost can be outweighed by the behavior of the tax base. The government's tradeoff between taxation and enforcement is strongly influenced by the degree of complementarity between firm size and the cost of tax avoidance. If greater firm size facilitates avoidance, then more competitive markets will feature lower tax avoidance rates, making higher tax rates rather than stiffer enforcement the lower-cost option for obtaining tax revenue.

1 Introduction

The idea that large, powerful corporations do not remit their fair share of taxes has gained considerable public traction in recent decades, both aided by and reflective of news reports of many of these top firms remitting little to no federal taxes. For example, in 2020 it was prominently reported that dozens of Fortune 500 companies remitted no federal income taxes the past year, with a subset of those also remitting nothing in the previous two years.[1] A 2023 Government Accountability Office (GAO) study reported that on average about a quarter of profitable large corporations remitted zero taxes between 2014 and 2018, and average effective tax rates hovered around 15% until dropping to single digits due to the Tax Cuts and Jobs Act of 2017 (TCJA).[2]

What fair means is subjective, further muddied by the distinction between legal avoidance and illegal evasion. Moreover, the majority, if not all, of these large firms remitting little to no taxes in a given year are reporting losses or carrying forward or back losses.[3] Yet, there is still a sense in many that total collections at the top end of the distribution should be higher. In a yearly Gallup poll taken between 2004 and 2019, the fraction of surveyed Americans who responded that corporations "pay too little" in taxes consistently stood at two-thirds, while the fraction believing corporations provide their "fair share" remained at one fifth.[4]

This paper examines efficiency based justifications for targeting tax and enforcement policy on market power. It develops a model of optimal business taxation and enforcement with imperfectly competitive industries. This model emphasizes two primary factors that drive the relationship between competition and tax policy. First, a more concentrated industry will already feature suboptimal production even in the absence of taxation. Increasing effective taxation will compound this market structure related distortion and therefore excess burden rises as concentration rises. Second, the elasticity of reported income encapsulates the fiscal behavioral response. As tax policy changes, so too will the tax base. This elasticity does not monotonically vary with respect to competition, which leads to the ambiguity in the relationship between tax policy and competition. However, regions of monotonicity can be established under certain assumptions.

The influence of the excess burden factor on optimal policy can be mitigated by increased cost deductibility. Under a pure profit tax in which all costs are fully deductible, production behavior is undistorted by the level of effective taxation and tax policy does not impose any excess burden regardless of market structure. While most corporate income taxes are profit taxes, there are reasons to believe that the effective fractions of cost deductibility fall below 100 percent. For example, no OECD country has a corporate tax system that features full loss offset, and many have limited depreciation allowances on capital expenditures.[5] Thus, this excess burden factor likely remains

[1] https://www.forbes.com/sites/tommybeer/2021/04/02/more-than-50-major-us-corporations-including-nike-and-fedex-paid-no-federal-taxes-last-year/?sh=3a7b338521d3

[2] https://www.gao.gov/products/gao-23-105384

[3] While carryback was disallowed by the TCJA 2017, the Coronavirus Aid, Relief, and Economic Security (CARES) Act in 2020 allowed firms to carryback losses 5 years for tax years beginning in 2018 to 2020.

[4] https://news.gallup.com/poll/1714/taxes.aspx

[5] The model in this paper does not directly incorporate loss years as it is static and firms must be weakly profitable to stay in business. But such ideas can be connected if the model is taken as a long-run perspective on profits.

relevant to some extent under real world profit taxes.

The elasticity of reported income can be decomposed into a true income elasticity and an avoidance elasticity. Unless there is a pure profit tax, the true income elasticity depends heavily on the demand elasticity. In the inelastic portion of demand, increasing competition decreases the income elasticity, which drives up optimal tax policy rates. In the elastic portion of demand, this relationship is mathematically ambiguous, and a positive relationship between competition and the income elasticity is possible.

The relationship between the avoidance elasticity and market power is intrinsically tied to the convexity of firm size, given by firm taxable income, in the avoidance cost function. If avoidance costs are concave in firm size, then firm size facilitates avoidance as it scales marginal benefits of avoidance more rapidly than the marginal costs of avoidance. In equilibrium, larger firms optimally choose higher avoidance rates. An increase in the tax rate would drive this up even further, while an increase in the enforcement more sharply reduces avoidance. All else equal, higher market power leads to larger firm sizes under mild assumptions in the symmetric Cournot setup of competition this paper employs.[6] Thus, the impact of the avoidance response on optimal taxation (enforcement) is negatively (positively) related to market power.

This size-avoidance cost complementarity then has direct implications for the government's tradeoff between statutory tax rates or enforcement. The core mechanical difference between these two tools is that taxes directly raise revenue while enforcement directly incurs real resource costs. The core behavioral difference is that taxes incentivize further avoidance while enforcement disincentivizes it.

Lower avoidance cost convexity in firm size implies both that larger firms will have higher levels of avoidance and are more responsive to both types of policy tools. This reduces the mechanical advantage of taxation while increasing the relative behavioral cost. This incentivizes a shift away from taxation toward heavier enforcement. Therefore, this lower firm size-avoidance cost convexity pushes toward a higher enforcement to tax ratio as concentration rises. Higher cost convexity implies the inverse. Only when firm size bears no impact on the avoidance decision is this tradeoff independent of market power.

The implication of these results are important for policy relevant discussion. Attitudes toward firm remittance behavior have not been exclusive to popular discourse. Auditing rates have consistently been correlated with firm size, with top firms having rates several factors higher than those of the smallest firms. As illustrated in Table 1, while there has been an overall decrease in enforcement rates across the board, the relative relationship between size and auditing probability has largely persisted. Though, this could simply be an illustration of an administrative cost advantage of auditing one very large firm versus auditing many smaller firms with an equal aggregate value, regardless of beliefs on underlying avoidance behavior. This administrative cost based argument is

[6]A symmetric Cournot competition setup establishes a direct link between firm size and market power within an industry. However, general comparisons between industries must have an "all else equal" modifier. Otherwise, differences in cost functions and market demand conditions can lead to a situation in which there are two equally sized firms in markets with strongly different market structures. This will be discussed in more detail in the paper.

explored in an extension.

Table 1: Auditing Rates by Firm Size (%)

Firm Size (Balance Sheet Assets)	2011	2013	2015	2017
\$1–\$999,999	1.00	0.81	0.55	0.32
\$1,000,000–\$99,999,999	2.48	2.03	1.94	1.45
\$100,000,000–\$999,999,999	20.50	17.31	16.15	9.41
> \$1,000,000,000	48.13	37.95	32.57	24.11

Data Source: IRS Data Book 2021, Table 17. Categories used by IRS redefined into the above four broader groups. More recent years are still potentially subject to change due to the 3-year rolling auditing process and so are not included.

More recently, there have been measures with more explicit targeting of top firms remitting more taxes. Two of the provisions of the landmark Inflation Reduction Act of 2022 in the US were to increase funding for the IRS by \$80 billion over the next decade, with \$45 billion earmarked for increasing and improving enforcement activities, and the introduction of a 15 percent Corporate Alternative Minimum Tax (CAMT) on corporations averaging over one billion dollars in income over a three year span. Among the plans for improvement was a stronger focus on tax system equity through enforcement on high income individuals and large firms who appear to remit considerably less taxes than expected.[7] This increase in funding was not a unanimously agreed upon decision, illustrated by an attempt by House Republicans to block the funding in January 2023, and later by a \$20 billion reduction in the funding as a part of the Fiscal Responsibility Act of 2023. From these conflicting decisions in a short time span, it is clear that there is little consensus on how enforcement should be handled and how it relates to the rest of the tax system.

The model in this paper does not directly feature an audit probability as illustrated above and as in the prototypical Allingham and Sandmo (1972) setup. Instead, a reduced form enforcement parameter enters into the firm's avoidance costs, similar to Kopczuk et al. (2016) and Keen and Slemrod (2017). The government can either spend money on enforcement to increase these costs for the firm or adjust the statutory tax rates. A firm in an industry with a fixed number of competitors then chooses output levels and avoidance rates to maximize profits subject to these policy tools.

As a starting point, I consider a tax system in which both of these tools are fully differentiable by industry, similar to the Ramsey (1927) commodity tax model. This allows for clearer illustrations of the forces at work. However, policymakers often face constraints on their instruments, and indeed most corporate tax rates are set uniformly. Thus this paper also analyzes a uniform tax system, but still allows for differentiated enforcement. While targeting suffers to some degree due to imperfect instruments, the general analysis carries through.

This paper connects to several strands of the literature. First it extends the optimal taxation literature that integrates agent evasion decisions and/or tax agency administrative decisions (Kaplow

[7]See introductory quote, from U.S. Secretary of the Treasury Janet Yellen's letter to the IRS: https://home.treasury.gov/system/files/136/JLY-letter-to-Commissioner-Rettig-Signed.pdf

(1990), Cremer and Gahvari (1993), Dharmapala et all. (2011), Keen and Slemrod (2017)). A particularly relevant point from Kaplow (1990) mirrors the previous discussion: despite similarities in the two types of tax tools, differences in behavioral distortions and resource costs imply a tradeoff and both can be used at the optimum.

I bridge these studies with the those that introduce market power into optimal tax problems (Besley (1989), Myles (1989), Kaplow (2021), Eckhout et al. (2021)). This area raises the relevance of considering the impact of firm entry/exit, tax rate pass-through, profit retention, and pre-existing output distortions in response to taxation. Pass-through in particular has received recent attention in how it relates to market power (Anderson et al. (2001), Weyl and Fabinger (2013), Pless and Van Benthem (2019), Miklós-Thal and Shaffer (2021), Adachi and Fabinger (2022), Ritz (2024)) and evasion (Kopczuk et al. (2016)).

Finally, this paper relates to the broad literature on the role of firms in the tax system (Kopczuk and Slemrod (2006), de Paula and Scheinkman (2010), Slemrod and Velayudhan (2018)). In much of the optimal tax literature, the direct role of firms is largely abstracted away from. However, in reality, firms remit the vast majority of taxes, and therefore being more direct in how their avoidance and enforcement may evolve is integral to understanding the tax system.

The empirical evidence for the connections between market power and avoidance rates is small, but has developed recent attention in the accounting literature. Kubick et al. (2015) find a negative association between product market power and effective tax rates, their proxy for tax avoidance, hypothesizing that persistent profitability due to market power allows firms to better predict future income streams and thus increases the value of tax avoidance strategies. Martin et al. (2022) examine the reverse direction and illustrate that firms that engage in tax avoidance have higher sales, aiding in higher concentration ratios. While these may not be enough to conclusively determine the true relationship between firm size and avoidance, they provide suggestive evidence.

The rest of this paper is structured as follows. Section 2 provides the details of the model for each type of agent: consumer, firm, and government. Section 3 focuses on how both taxation and enforcement affect the firms in the markets, in their profits, and in how their output and avoidance decisions affect reported taxable incomes. Section 4 introduces and discusses optimal policy expressions, both for the levels and in the tradeoff between the two policy tools. Section 5 illustrates features of the model and examine these interactions through a simulated environment. Section 6 constricts the fully differentiated tax rates to a uniform system and examines how the intuition from the Ramsey setup carries over. Section 7 covers a few extensions and alterations of the model. Section 8 discusses the main results of the paper and concludes. The majority of the mathematical derivations are relegated to the Appendix.

2 Model Setup

There is a set of K industries with potentially varying degrees of competition in the economy. Though a perfectly competitive or monopoly industry may not exist in this set, the results of these two extreme cases may sometimes be discussed to help build intuition. Each of these industries produces a unique good k such that the terms good and industry are used interchangeably when referring to the differentiated tax or enforcement rates (e.g., a "tax on good k" or a "tax on industry k"). For goods in each industry, consumers face a tax rate inclusive price q_k while producers receive the net of tax price of $p_k = q_k(1-t_k)$. It may also be that industries can be partitioned into isolated markets. In this case k can instead be treated as a good-market combination.

2.1 Consumer Problem

A representative consumer chooses an optimal consumption bundle x of the K goods in addition to leisure (or, equivalently, choosing labor L). Labor is treated as the numeraire good such that the wage is set to 1. The consumer receives all profits Π_k from each industry. Thus, the consumer solves

$$\begin{aligned} &\max_{\vec{x},L} u(\vec{x}, L) \\ &\text{s.t. } I + L + \sum_k \Pi_k \geq \vec{q} \cdot \vec{x} \end{aligned} \tag{1}$$

While I denotes additional nonlabor income, this will be set to 0 and its purpose here is to define the marginal utility of income, $\frac{\partial v}{\partial I} = \alpha$, where v is the indirect utility. The individual does not consider their own impact on either prices or on profits. The model assumes away income effects in the demand for each good via quasilinear utility and cross-price effects across markets to simplify exposition of the key results.

2.2 Firm Problem

A firm in industry k chooses output y_k and an avoidance rate γ_k, while subject to a tax rate t_k and an enforcement rate β_k, to maximize their profits. While the term avoidance will be used throughout, the model makes no distinction here between illegal evasion and legal avoidance, and γ_k serves to capture both of these types of actions. One limitation with this assumption is that marginal costs for these behaviors may not ever align in practice. Thus, using a single variable ignores a potentially important decision margin in how the firm decides to reduce tax remittance and in how the government decides to limit these behaviors.

Firms have producer price $p_k(x_k) = q_k(x_k)(1-t_k)$ and are allowed to deduct a fraction $\mu \in [0, 1]$ of their direct production costs $\phi_k(y_k)$ from their taxable income. Firm size is defined to be pre-avoidance taxable income: $r_k \equiv y_k q_k - \mu\phi_k(y_k)$. Avoidance costs $H_k(r_k, \gamma_k, \beta_k)$ are a multiplicatively separable function of firm size, the avoidance rate, and the enforcement rate. Total costs are a sum of these direct production costs and avoidance costs: $C_k(y_k, r_k, \gamma_k, \beta_k) = \phi_k(y_k) + H_k(r_k, \gamma_k, \beta_k)$. From this point on, the k subscript in the cost function will be dropped in favor of partial derivative representation, e.g., $H_\gamma \equiv \frac{\partial H(r_k,\gamma_k,\beta_k)}{\partial \gamma_k}$, but it should be remembered that C may functionally differ across industries.

In order to obtain feasible and interior solutions for output and avoidance, ϕ and H are assumed to be continuous and twice differentiable in all arguments and $\phi_y > 0$ for $y \geq 0$, $H_\gamma > 0$ for $\gamma \geq 0$, and $H_{\gamma\gamma} > 0$. As increasing enforcement should increase the costs of avoidance, it is also assumed that $H_\beta > 0$ for $\gamma > 0$ and $H_{\gamma\beta} > 0$. The case of decreasing marginal production costs, for which a less competitive industry may be socially desirable, will be set aside. Formally, the firm solves

$$\max_{y_k, \gamma_k} y_k q_k(x_k) - C_k(y_k, r_k(y_k), \gamma_k, \beta_k) - t_k(1-\gamma_k) r_k(y_k) \tag{2}$$

If $\mu = 0$, no costs are deductible and firms face a pure output tax. If $\mu = 1$, all production costs are deductibile and firms face a pure profit tax. This affects the relevant tax base for the firm–they hide revenues under an output tax and production profits under a profit tax. The relevant tax base is what enters as size into the firm's cost function with the implicit assumption that the exact functional form for avoidance costs may differ whether the firm is hiding revenues or profits.

Another note here is that avoidance costs are not included as deductible costs for the firm or as part of the portion of income that the firm avoids. Most avoidance costs in practice are likely to be cost deductible to the firm. However, including deductibile avoidance costs does little to change the results while significantly complicating the exposition of the expressions. Thus, the version of the model with deductible avoidance costs version is left to the Appendix.

Imperfectly competitive markets have a fixed number N_k of identical Cournot competitors. This Cournot framework has a few implications. First, it implies that within a market, number of firms, firm size, and market power are isomorphic. Between markets, this is not true. It is possible for equivalently sized firms in two industries to face very different market structures. Thus, most results discussing the impact of competition are framed as either a "within market" or an equivalent all else equal "between market" approach. The thought experiment in the former approach is to exogenously increase (or decrease) the number of firms in a given market. The latter is to consider two markets that are equivalent in all fundamentals except in the number of competitors.[8]

The firm's profit maximizing production condition is

$$q_k \left[1 - \frac{1}{N_k \varepsilon_{x_k}}\right] = \frac{(1 - \mu[(1-\gamma_k)t_k + H_r])}{(1 - (1-\gamma_k)t_k - H_r)} \phi_y \tag{3}$$

where $\varepsilon_{x_k} \equiv -\frac{\partial x_k}{\partial q_k}\frac{q_k}{x_k}$ is the positively defined elasticity of demand. This is the typical equating of marginal revenues and (effective) marginal costs. Under a pure profit tax ($\mu = 1$) the solution for the optimal choice of output is independent of the tax rate and the avoidance rate. In other words, as is the case in models that do not feature avoidance, the production choice is undistorted by the tax and the equilibrium quantity is the same as if there were no tax.

This expression implies that

$$1 - \frac{1}{N_k \varepsilon_{x_k}} > 0$$

[8] An alternative between market comparison would be that in Ritz (2024). In this case, "all else equal" refers to observed equilibrium values, e.g., the observed equilibrium value of the demand elasticity. For the demand elasticity of two markets with different market structures to be equal, it must be the case that the fundamental demand curve is different.

for nonzero marginal costs. One important takeaway from this formulation is that firms in the industry must be producing in the portion of the demand curve where $1 - \frac{1}{N_k \varepsilon_{x_k}} > 0$, or $\varepsilon_{x_k} > \frac{1}{N_k}$. A commonly used version of this result is that a monopoly must produce in the elastic part of the demand curve, i.e., where $\varepsilon_{x_k} > 1$. This conclusion, and the more general Cournot conclusion, still hold with avoidance. As N_k increases, this lower bound on the allowable portions of the demand curve for production falls.

Lastly, the elasticity-adjusted Lerner index (LHS) can be written as

$$\frac{q_k - \frac{1-\mu[(1-\gamma_k)t_k + H_r]}{1-(1-\gamma_k)t_k - H_r}\phi_y}{q_k}\varepsilon_{x_k} = \frac{1}{N_k} \tag{4}$$

where the equality follows from the production condition. The Lerner index represents the adjusted price to marginal cost ratio, which gives a sense of the market power in an industry. In a Cournot framework this is exactly related to the (inverse) of the number of firms, which warrants the use of number of firms as a measure of competition.

The profit maximizing avoidance condition is given by

$$t_k r_k = H_\gamma(r_k, \gamma_k, \beta_k) \tag{5}$$

which follows the same intuition of equalizing marginal benefits and costs. The marginal benefit of avoidance, the LHS of the FOC, increases as firm size increases in a linear manner. If the marginal costs of avoidance also rise linearly with firm size, such that $H(r_k, \gamma_k, \beta_k) = r_k H(1, \gamma_k, \beta_k)$, then the avoidance rate is independent of firm size and correspondingly the concentration of firms. If instead $H(\cdot)$ is convex in firm size, then the optimal avoidance rate is inversely related to firm size. Since per-firm size of an industry decreases as the number of firm increases, for a given tax-enforcement combination the avoidance rate increases as competition increases.[9] The intuition is that if avoidance costs are convex in size, then growing larger hinders a firm's ability to avoid as costs outpace benefits. The reverse is true if the cost of avoidance is concave in size.

The degree of convexity matters here. As convexity (concavity) is amplified, so are the directionalities of the above statements. In order to get at this idea, define the elasticity of the marginal costs of avoidance with respect to firm size as

$$\varepsilon^r_{H_\gamma} \equiv H_{\gamma r}\frac{r_k}{H_\gamma}$$

An elasticity of 1 represents marginal avoidance costs that scale linearly with size, a special subcase which will be referred to as constant returns to scale avoidance technology. An elasticity greater than 1 represents the case where marginal avoidance costs are convex in size, decreasing returns to scale avoidance technology, and less than 1 represents the case where costs are concave in size, increasing returns to scale avoidance technology. The importance of this relationship is summarized

[9] Firm size decreasing with the number of firms is shown in the appendix.

in the following Lemma.

Lemma 1. *As the elasticity of marginal avoidance costs with respect to firm size, $\varepsilon^{r}_{H_\gamma}$, increases, the more positive is the within industry relationship between avoidance rate and competition. For values of the elasticity less than 1, the relationship is negative. For values of the elasticity greater than 1, the relationship is positive. At an elasticity of 1, there is no relationship between the number of firms and the avoidance rate.*

It should be reiterated that avoidance is discussed throughout this paper as a rate. If the rate of avoidance is the same for two differently sized firms, then the larger firm must be avoiding more on a level basis. Even under decreasing returns to scale technology, a larger firm could be avoiding more on a level basis.

Importantly, this elasticity does not directly imply anything about the relationship between firm size and avoidance rates across markets unless we assume that the same avoidance technology is available to all firms regardless of industry, an assumption highly unlikely to have any basis in the real world. This would be require that there are neglible industry-specific benefits or access to avoidance opportunities (e.g., specialized accountants in one industry that have more expertise than those in another industry).

2.3 Government

The government seeks to maximize the utility of the representative agent subject to the use of two sets of policy instruments, industry specific tax rates t and industry specific enforcement efforts β in order to satisfy the following budget constraint:

$$\sum_j t_j(1-\gamma_j)R_j - a(R_j,\beta_j) \geq G \tag{6}$$

where R_j is aggregate industry size (pre-avoidance taxable production income), $a(\cdot)$ is the administrative cost function, and G is an exogenous revenue requirement that will ultimately determine the value of the government budget multiplier λ.[10] Administrative costs are a function of both the effort put in by the government into the industry, β_j, as well as the size of the industry. While sometimes separated out into real production income and avoidance components, define total reported taxable income of an industry $Z_j \equiv (1-\gamma_j)R_j$. The Lagrangian of the full problem is then

$$\mathcal{L} = v(q) + \lambda\left[\sum_j t_j Z_j - a(R_j,\beta_j) - G\right] \tag{7}$$

[10] A note on the value of λ when performing "within market" comparisons. In the hypothetical comparison we often perform, which is to alter the number of firms (and therefore market power) within an industry, we are implicitly also structurally changing the overall economy as the equilibrium aggregates absent taxation are altered. In this sense, λ is endogenous to the degree of market power we attribute to a given industry. From this point of view, an "all else equal" between market comparison may be preferable. Otherwise, we can assume that that the relative size of the rest of the economy is large enough such that structural changes in one industry neglibly affect the value of λ.

The cost of enforcement is normalized to one such that administrative costs can take the form $a(R_j, \beta_j) = R_j\beta_j$. In this form, β_j can directly be interpreted as a rate similar to the tax rate t_j. The expression $t_j(1 - \gamma_j) - \beta_j$ can be defined as the mechanical net rate of revenue raised by the government for a given policy combination.

The enforcement parameter does not directly incorporate fixed penalty fees on the firm side or fixed administrative cost on the government side. Incorporating the former would imply an avoidance "entry" decision for the firm, where firms may choose not to avoid at all if the profit gain from their optimal avoidance choice falls below the (expected) fee paid. Incorporating the latter would have a similar extensive margin for the government. The government would choose to ignore the smallest firms for which enforcement activity would not have a net positive return.

The fully differentiated system above will be useful for deriving intuition. However, most corporate tax rates observed in the world are uniform. Thus, in section 6, I consider an alternative model where the tax is constrained to be uniform across all industries, though enforcement rates can still be differentiated. This is in line with the assumption that enforcement authorities such as the IRS have relatively more freedom in how they choose to target their resources whereas corporate tax rates are much more strongly impacted by political economy constraints.

3 Market Responses to Taxation and Enforcement

Important to the discussion of the optimal tax and enforcement rates are how firms and therefore the markets respond to these tools. As is typical in optimal tax problems, the two primary components to balance are the elasticity of the base of the tax tool and the welfare impacts on consumers. The welfare impact of a policy change is composed of the direct price effects and profit responses, since consumers receive all profits from each industry. Behavioral impacts of bundle re-optimization are zeroed out by the envelope condition and income and cross-price effects are assumed zero. Thus, the welfare effect is defined by the expression

$$\frac{dv}{d\theta_k} = -\frac{\partial q_k}{\partial \theta_k} x_k + \frac{\partial \Pi_k}{\partial \theta_k} \tag{8}$$

where v is the indirect utility of the consumer and $\theta_k \in \{t_k, \beta_k\}$ represents one of the tax tools. A relevant exercise that will help determine the tradeoff between the use of taxation and enforcement is an offsetting change in one tool versus the other. One such exercise is to offset the consumer price change (or equivalently the output change) by adjusting the second tool to just offset the output distortion. In this case, the net welfare effect is

$$\frac{dv}{dt_k} - \frac{dv}{d\beta_k}\frac{\frac{\partial q_k}{\partial t_k}}{\frac{\partial q_k}{\partial \beta_k}} = \frac{\partial \Pi_k}{\partial t_k} - \frac{\partial \Pi_k}{\partial \beta_k}\frac{\frac{\partial q_k}{\partial t_k}}{\frac{\partial q_k}{\partial \beta_k}} \tag{9}$$

On the fiscal side, the base of the tax rates are the reported taxable incomes in each industry Z_k. Define the elasticities of reported income with respect to each of the tax rates and enforcement rates respectively as

$$\varepsilon_{Z_k} \equiv \frac{\partial Z_k}{\partial(1-t_k)}\frac{1-t_k}{Z_k}, \qquad \varepsilon^{\beta}_{Z_k} \equiv \frac{\partial Z_k}{\partial \beta_k}\frac{\beta_k}{Z_k}$$

These will be the two relevant elasticities in determining an inverse elasticity type rule, similar to the demand elasticity in the standard Ramsey commodity tax framework and the elasticity of taxable income in an indvidual income tax framework. Before the full discussion of how these elasticities and consumer welfare are affected by the presence of avoidance and market power, we will first discuss in detail the individual components: price responses, avoidance responses, and profit responses.

3.1 Price Response

The pass-through rate of each tax tool onto consumer prices is important for both welfare and reported income behavior. In a standard Ramsey framework, free entry and constant returns production technology in perfectly competitive markets ensure that the entire burden of the tax is shifted to the consumer, i.e., $\frac{\partial q_k}{\partial t_k} = 1$ if t_k is an excise tax and $\frac{\partial q_k}{\partial t_k} = \frac{q_k}{1-t_k}$ if t_k is ad valorem, and thus the price change is always the same no matter where the tax is applied (on a level basis for excise taxes and percentage basis for ad valorem). The introduction of both avoidance and market power into the problem alters this idea.

With avoidance, firms do not bear the full burden of taxation and thus they also do not pass the full burden of taxation onto consumers either. Rather, they pass on the portion of the tax burden they do experience, ie., the effective tax.[11]

Likewise, the lack of free entry in imperfectly competitive markets and the (partial) control of pricing implies that pass-through can fall below the competitive level or can even be overshifted. As discussed in depth in Weyl and Fabinger (2013), the degree of pass-through in imperfect markets hinges on the concavity of the demand function. Defining the negative of marginal consumer surplus $ms_k = -\frac{\partial q_k}{\partial x_k}x_k$, then log-concave demand functions imply $\frac{1}{\varepsilon_{ms_k}} > 0$, which ensures that pass-through is below the competitive level for a monopoly and rises as competition increases so long as costs are not too convex (Ritz, 2024).[12]

Combining these two considerations, define and derive the tax pass-through rate as

$$\rho_k \equiv \frac{1}{q_k}\frac{\partial q_k}{\partial t_k} = \frac{\left(1-\gamma_k - t_k\frac{\partial \gamma_k}{\partial t_k}[1-\varepsilon^r_{H_\gamma}]\right)\left[1-\frac{1}{N_k\varepsilon_{x_k}} - \mu\frac{\phi_y}{q_k}\right]}{(1-t_k+\gamma_k t_k)\left(1+\frac{1}{N_k\varepsilon_{ms_k}} + \frac{\varepsilon_{x_k}-\frac{1}{N_k}}{\varepsilon_{s_k}}\right)} \tag{10}$$

[11] In a problem with avoidance costs as specified in our model, the "effective tax" faced by the firm is not the same as the "effective tax" from the point of view of the government. The latter is the actual rate of tax collection for a given statutory tax rate and is given by $(1-\gamma_k)t_k$. The prior includes includes additional changes in net margins due to altering firm size and is given by $(1-\gamma_k)t_k + H_r$.

[12] Conditions for demand and cost convexity are given by Proposition 2 in Ritz (2024). "Not too convex" means that the $\eta_s \equiv \frac{1}{\varepsilon_s} \in (0, 1]$.

and the enforcement pass-through rate as

$$\rho_k^\beta \equiv \frac{1}{q_k}\frac{\partial q_k}{\partial \beta_k} = \frac{\left(H_{r\beta} - t_k \frac{\partial \gamma_k}{\partial \beta_k}[1-\varepsilon^r_{H_\gamma}]\right)\left[1 - \frac{1}{N_k \varepsilon_{x_k}} - \mu\frac{\phi_y}{q_k}\right]}{(1 - t_k + \gamma_k t_k)\left(1 + \frac{1}{N_k \varepsilon_{ms_k}} + \frac{\varepsilon_{x_k} - \frac{1}{N_k}}{\varepsilon_{s_k}}\right)} \tag{11}$$

where ε_{x_k} is the elasticity of demand and ε_{s_k} is the elasticity of the inverse marginal cost (hereby "supply") curve. The derivation of these pass-through rates follows closely to Weyl and Fabinger (2013) and Adachi and Fabinger (2022), with modifications associated with the inclusion of avoidance.

One key term that appears in both expressions is the effect of the endogenous change in the avoidance rate: $-\frac{\partial \gamma_k}{\partial \theta_k}[1 - \varepsilon^r_{H_\gamma}]$. The term in the brackets respresents how the net margins of avoidance are related to firm size. If this is positive, additional avoidance by the firm will increase the net margins of the firm. The industry passes a portion of this net benefit onto the consumer through a lower pass-through rate. Since $\frac{\partial \gamma_k}{\partial t_k} > 0$, this is exactly what happens with the tax rate. For the enforcement rate, $\frac{\partial \gamma_k}{\partial \beta_k} < 0$, and so the firm will incur net costs from changing avoidance and pass on those costs to the individual instead via a higher pass-through rate. All effects are reversed if $\varepsilon^r_{H_\gamma} > 1$ and firm size is negatively related to the net margins of avoidance.

A final difference in these expressions with the avoidance-less versions is in the supply elasticity. The marginal costs defined in this term are inclusive of how additional production affects avoidance costs. Therefore, even if there are constant marginal production costs, which would typically imply $\varepsilon_{s_k} = \infty$ and this overall term would vanish, this elasticity would remain.

For constant returns to scale avoidance technology, the endogenous avoidance effect drops out of both expressions. More importantly, the qualitative relationship between these expressions and market power are the same as if there were no avoidance. One useful implication of this is that for log-concave demand functions, the pass-through rate for both tools is unambiuously increasing as competition increases. This point is repeated in the following lemma.

Lemma 2. *Under constant scale avoidance technology, if a market faces a log-concave demand function and the firm production costs are not too convex, the pass-through rate of either policy tool increases as the number of firms, and therefore competition, increases.*

3.2 Avoidance Response

Section 2 discussed the implications of the elasticity of marginal avoidance costs on the levels of avoidance. This section now discusses the implications on the behavioral response of avoidance, one of the two components of the reported taxable income response. The change in the avoidance rate in response to an increase in the tax rate is

$$\frac{1}{1-\gamma_k}\frac{\partial \gamma_k}{\partial t_k} = \frac{1}{(1-\gamma_k)H_{\gamma\gamma}}\left[\underbrace{r_k}_{Direct} + \underbrace{t_k(1-\varepsilon^r_{H_\gamma})\frac{\partial r_k}{\partial t_k}}_{\text{Size Response}}\right] \tag{12}$$

which decomposes the response into a direct effect and an indirect size related effect. The first effect is the direct consequence of the tax increase making avoidance more attractive as the marginal benefit of avoidance increases. The latter is due to the firm reoptimizing its production and therefore size in response to the tax change.

The term $1-\varepsilon^r_{H_\gamma}$, as before, determines whether this change in firm size is beneficial or not. Again, this term represents the change in the net margins of avoidance as firm size increases. Therefore, $1-\varepsilon^r_{H_\gamma} > 0$ implies that increasing firm size increases the net margin of avoidance, making further avoidance more attractive.

The direction of the change in firm size is determined by the determined by the value of the demand elasticity. For inelastic demand, reductions in aggregate output increase total industry revenue. Since number of firms in constant, this implies a higher per-firm revenue. For elastic demand, the reverse is true and the per-firm revenue decreases.

Combining these terms obtains the sign of the overall size response term. If $\varepsilon_{H^r_\gamma} < 1$, then a demand elasticity greater than 1 (under an output tax such that $\mu = 0$) indicates a negative overall response while an elasticity lower than 1 indicates a positive response. These directions are inverted if $\varepsilon^r_{H_\gamma} > 1$. Thus, the endogenous size response may either amplify or mute the direct response depending on where along the demand curve the firms produce in.

Constant scale avoidance implies $1-\varepsilon^r_{H_\gamma} = 0$, leaving only the direct effect. Since firm size is unrelated to the avoidance decision, the endogenous change in firm size has no further impact on attractiveness of avoidance. This assumption also implies that the direct effect is unrelated to firm size since

$$\frac{r_k}{(1-\gamma_k)H_{\gamma\gamma}(r_k,\gamma_k,\beta_k)} = \frac{1}{(1-\gamma_k)H_{\gamma\gamma}(1,\gamma_k,\beta_k)}$$

under constant scale avoidance. Since γ_k does not change with firm size, the direct effect is independent of firm size and market power. Under increasing scale avoidance technology, however, the direct effect would increase with market power. Since the marginal benefit of avoidance scales faster with size than do marginal costs, larger firms, conditional on current firm size, will want to engage in relatively more avoidance.

The corresponding expression for the change in the enforcement rate is given by

$$\frac{1}{1-\gamma_k}\frac{\partial \gamma_k}{\partial \beta_k} = \underbrace{-\frac{H_{\gamma\beta}}{(1-\gamma_k)H_{\gamma\gamma}}}_{Direct} + \underbrace{t_k(1-\varepsilon^r_{H_\gamma})\frac{\partial r_k}{\partial \beta_k}}_{\text{Size Response}} \tag{13}$$

A similar discussion to the tax counterpart can be had here. The key difference is that the direct effect is to incentivize lower avoidance as enforcement directly increases the marginal cost of avoidance. Thus, while the endogenous size response may amplify or mute the direct effect here as well, it works in the opposite direction as in the tax case.

Similarly, the relationship between the direct effect of an enforcement change and market power is generally reverse to that of a tax change (both increase in magnitude, but in opposite signs). Under increasing returns avoidance technology, larger firms engage in more avoidance for a given enforcement and tax rate. Increasing the enforcement rate directly impacts the marginal avoidance costs of the firm. Since avoidance costs are strictly convex in the avoidance rate, this implies a larger impact at higher levels of avoidance.

The following lemma summarizes the relationships between tax and enforcement with market power.

Lemma 3. *Under constant returns to scale avoidance technology, there is no relationship between market power and the avoidance response for a given industry. Under increasing (decreasing) returns to scale avoidance technology, there is a positive (negative) relationship between market power and the direct effect of taxes on avoidance and a negative (positive) relationship between market power and the direct effect of enforcement on avoidance. Under a profit tax,* $\mu = 1$, *this direct effect represents the entire avoidance response.*

3.3 Profit Response

The final building block to discuss are the profit responses to the tax tools. The aggregrate profit response for an industry k is

$$\frac{\partial \Pi_k}{\partial t_k} = \underbrace{-Z_k}_{Mechanical} + \underbrace{(1-(1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\rho_k x_k q_k}_{Competition} \tag{14}$$

The first term represents the mechanical change in profits for the industry, which is the negative of aggregate reported income. Under a monopoly ($N_k = 1$), any behavioral impact is zeroed out by the envelope condition. With $N_k > 1$ competitors, each firm in the market has some control over prices and a joint behavioral impact still arises. This effect rises with competition and approaches the negative of the mechanical effect. Thus, the overall response trends to zero as should be expected in competitive markets.

Similarly, the impact of a change in the enforcement rate on aggregate profits is

$$\frac{\partial \Pi_k}{\partial \beta_k} = -N_k H_\beta + (1-(1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\rho_k^\beta x_k q_k \tag{15}$$

where we see the same separation into the mechanical effect of the increase in enforcement rates on profits and the competitive adjustment due to the partial price control each firm has. The direct effect here is the marginal increase in avoidance costs for the firms with respect to enforcement, sometimes referred to as the marginal compliance cost.

3.4 Welfare Effect

The total welfare effect is the sum of the price effect on consumers and the profit change. Using the profit response expression, the welfare effect in response to a tax change can be written as

$$\frac{1}{\alpha}\frac{dv}{dt_k} = -Z_k - \rho_k x_k q_k \left[1 - (1 - (1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\right] \tag{16}$$

Breaking down this expression, the first term is the reported income in the industry, which exactly equals the mechanical gain in tax revenue. The second term represents the additional welfare effect beyond the mechanical transfer in tax revenue from firms/consumers to the government.

Dividing through by reported income converts this expression to the welfare effect per mechanical dollar raised:

$$\frac{1}{\alpha Z_k}\frac{dv}{dt_k} = -1 + EB_k \tag{17}$$

where the second term has been redefined as the marginal excess burden of taxation per dollar raised, labeled EB_k.

$$EB_k = \rho_k \frac{x_k q_k}{Z_k}\left[1 - (1 - (1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\right] \tag{18}$$

One thing to note is that the excess burden term does not trend exactly to zero even in perfect competition. The importance of the term, however, is that it grows as market power increases and therefore is indicative of the increasing distortion in the market.

Similarly, combining the results to obtain the welfare effect per reported income on the enforcement side gives

$$\frac{1}{\alpha Z_k}\frac{dv}{d\beta_k} = -\frac{N_k H_\beta}{Z_k} - EB_k^\beta \tag{19}$$

where EB_k^β is equivalent to EB_k except in replacing the tax pass-through with the enforcement pass-through. While this expression does not have as natural of a breakdown into an expected transfer of a dollar from the consumers or firms to the government and the excess cost, similar logic can be applied. The first term is the impact on profits per dollar of mechanical tax revenue and is thus the "standard" burden. The second term is the welfare loss above and beyond this mechanical impact.

Finally, this excess burden effect can be described in terms of potentially observable characteristics. Under the profit tax the excess burden is zero, so we focus on the output tax case here. Using the production condition, the excess burden can be written as

$$EB_k^\theta = \rho_k^\theta \left(1 - \frac{\frac{\phi_y}{q_k}}{1 - \frac{1}{N_k \varepsilon_{x_k}}}\right)\left(1 - \frac{1}{N_k}\right) \tag{20}$$

Within a market, the excess burden is increasing with market power, but this expression can also be used compare market power across industries. If two industries have similar pass-through rates and demand elasticities, the market with higher market power (lower number of firms) has the higher excess burden. All else equal, this will provide a stronger downward force on both the optimal tax and enforecement levels, as will be seen in Section 4.

The most challenging portion of this term to empirically observe is the marginal cost of production. These has been a increasing trend to back these costs out in empirical industrial organization research. A crude alternative is to assume constant marginal costs and no fixed costs. In this case, marginal costs are equivalent to average costs. This can then be calculated as total costs, which are often reported, divided by total quantity.

The last term in the welfare expressions thus far not described in terms of potential observables is the marginal complicance cost to the firm H_β. There has been recent work uncovering compliance costs, both directly related to administration and enforcement (Harju et al. (2019)) or related to other regulatory compliance (Trebbi et al. (2023)).

3.5 Reported Income

The final step is to describe the behavior of the tax base of reported industry income $Z_k \equiv (1-\gamma_k)R_k$. The elasticity of reported income to the retention rate is

$$\frac{\varepsilon_{Z_k}}{1-t_k} = \underbrace{\frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{1}{1-\gamma_k}}_{\text{Avoidance Response}} - \underbrace{\frac{(1-\gamma_k)x_k q_k}{Z_k}\rho_k\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]}_{\text{True Taxable Income Response}} \tag{21}$$

which can be decomposed into the avoidance response and the true taxable income (or industry size) response. The avoidance response has been detailed in Section 3.2. In the avoidance response, there is a firm size response that can be joined to the industry size response here but these will be kept separate for now.

The corresponding change in the reported income to a change in the enforcement rate is

$$\frac{\varepsilon_{Z_k}^{\beta_k}}{\beta_k} = \frac{\partial(1-\gamma_k)}{\partial\beta_k}\frac{1}{1-\gamma_k} + \frac{(1-\gamma_k)x_k q_k}{Z_k}\rho_k^{\beta}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right] \tag{22}$$

To simplify exposition, the two poles of cost deductibility, profit taxation ($\mu = 1$) and output taxation ($\mu = 0$), will be discussed in turn.

3.5.1 Profit Taxation

Under profit taxation, the true taxable income response is zeroed out since the production decision is unaffected by the effective tax rate. This implies that the reported income response is equivalent to just the avoidance response. Thus, the relationship between the reported income response and market power exactly follows the relationship between avoidance respose and market power, as described in Lemma 3. Under constant scale avoidance technology, there is no relationship. Under increasing scale technology, the magnitude of the response for both tools increases as market power increases.

3.5.2 Output Taxation

Under an output tax, the true income response does not drop out, but the overall expression can be simplified to

$$\frac{\varepsilon_{Z_k}}{1-t_k} = \frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{1}{1-\gamma_k} - \rho_k\left[1-\varepsilon_{x_k}\right] \tag{23}$$

and a similar expression for the enforcement elasticity. Note that the taxable income here is just reported revenue. For inelastic demand, $\varepsilon_{x_k} < 1$, an increase in the tax (enforcement) rate decreases (increase) the reported income elasticity. This is because an increase in the effective tax rate increases industry revenue under inelastic demand. This broadening of the tax base is a beneficial fiscal effect and thus pushes the elasticity in the respective beneficial directions for each tool.

More important is how these expressions relate to market power. Consider the constant avoidance technology case. The avoidance response is independent of firm size. Then the only thing that matters is how the pass-through rate and the demand elasticity change with number of firms, i.e., how the elasticity of true revenue changes with the number of firms. As discussed previously, it can be shown that log-concave demand functions imply increasing pass-through rates with competition. We can then make the following assertions.

Proposition 1. *Under constant scale avoidance technology, the relationship between market power and the elasticity of reported revenue with respect to the relevant tax tool is qualitatively the same as the relationship between market power and elasticity of true revenue with respect to the tool. For log-concave demand functions and production costs that are not too convex, an increase in competition decreases (increases) the elasticity of reported revenue to taxation (enforcement) if* $\varepsilon_{x_k} < 1$.

For elastic demand, the relationship between market power and the size of the true income response is ambiguous. Therefore, unit elasticity represents a sort of "tipping point" between this ambiguous region and the relationship described in the previous proposition. Under non-constant returns avoidance techonology, this tipping point changes as the avoidance response shifts where the change from a positive to a negative industry size effect on the overall response occurs.

4 Optimal Rates and Tradeoff

Optimal policy rates and the tradeoff between the two tools at the government's disposal are now derived. Much of these results will simply be formalizations of moving parts decribed in the previous section.

4.1 Optimal Tax Rates and Enforcement Rates

Tying everything together, the expression for the optimal tax and enforcement rates in the presence of market power and avoidance are given by the following proposition:

Proposition 2. *Given an enforcement rate, the optimal tax rate in an industry k is*

$$\frac{t_k}{1-t_k} = \frac{1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right] - \frac{\beta_k}{Z_k}\frac{\partial R_k}{\partial t_k}}{\varepsilon_{Z_k}} \tag{24}$$

Given a tax rate, the optimal enforcement rate in an industry k is

$$\beta_k = \frac{t_k \varepsilon^{\beta}_{Z_k}}{\frac{R_k}{Z_k} + \frac{\beta_k}{Z_k}\frac{\partial R_k}{\partial \beta_k} + \frac{\alpha}{\lambda}\left[\frac{N_k H_\beta}{Z_k} + EB^{\beta}_k\right]} \tag{25}$$

Again EB_k is representative of the additional distortion taxation has due to the markets being imperfectly competitive. As the number of firms rises, excess burden decreases, providing a driving force upward on the tax rate as competition rises (alternatively stated, as market power increases this provides a downward force on the tax rate). This is true for both policy tools as either will increase the effective tax rate of the firm, which what the firm responds to and distorts the market.

However, as derived in the previous section, the behavior of reported income does not necessarily follow the same direction. Since the reported taxable income elasticity is in the denominator of the tax expression, a lower response drives up the tax rate while a higher response drives down the tax rate. Since the excess burden term decreases with competition, if the reported income elasticity also decreases with competition, then we can unambiguously say that the optimal tax rate should increase with competition. As discussed in Proposition 1, one case where this is true is under constant scale avoidance technology, log-concave demand, and an elasticity of demand below unity (under an output tax).

While we cannot guarantee that for the same conditions, but an elasticity greater than 1, that increasing market power should increase the optimal tax rate, we can say that elasticity greater than 1 is a necessary additional condition for this relationship to be possible. Relaxing the constant scale avoidance technology will alter the benchmark elasticity that separates the first region from the second. Simulations in Section 5 will illustrate this possibility for optimal tax (and enforcement) rates to be increasing with market power in certain regions.

The optimal enforcement expression follows much of the same intuition. The imperfect market related distortion is positively in the denominator, indicating that a higher value decreases the optimal enforcement rate, as should be expected. The enforcement elasticity is in the numerator, indicating a higher value increases the optimal enforcement rate. This too makes sense as a higher enforcement elasticity implies that enforcement is highly effective in increasing reported income which allows the tax rate to be more effective. As discussed in Proposition 1, the same conditions that allow for the elasticity of taxable revenue and the market distortion factor to go in the same direction also allow the corresponding two effects to go in the same direction for the enforcement expression.

Not discussed yet are the enforcement costs in each expression. For the tax rate, there are no direct enforcement costs, but there is an indirect effect since enforcement costs are based on industry size. Since a change in the tax rate alters the (true) firm and industry size, it will change how much is spent on total enforcement. If the industry grows, then this leads to higher total enforcement costs, and so this puts a downward effect on the tax rate. This same indirect effect also occurs in the enforcement expression, and similarly an increase in industry size would be a negative factor in the enforcement rate and drive it down. The enforcement rate also has a "mechanical" resource cost, but this is simply the current industry size. If administrative costs are combined with the income elasticity, we can obtain an "enforcement elasticity of tax revenue" in a vein identical to Keen and Slemrod (2017).

Under a pure profit tax, these relationships become even simpler. There is no excess burden related to production and market power, leaving only the reported taxable income elasticity in the tax expression. For enforcement, the excess burden also drops out out, but there is still an administrative cost consideration.

Corollary 2.1. *Under a profit tax, the conditional optimal tax and enforcement rates are given by*

$$\frac{t_k}{1-t_k} = \frac{1-\frac{\alpha}{\lambda}}{\varepsilon_{\gamma_k}}, \qquad \beta_k = \frac{t_k \varepsilon^{\beta}_{\gamma_k}}{\frac{R_k}{Z_k} + \frac{\alpha}{\lambda}\frac{H_\beta}{z_k}} \tag{26}$$

Under constant scale avoidance technology, the levels of both the profit tax and enforcement rate are independent of market power. Under increasing (decreasing) returns to scale avoidance technology, the tax (enforcement) is negatively (positively) related to market power all else equal.

Though this corollary replaces the elasticity of reported taxable income with the elasticity of avoidance to highlight the lack of a production distortion under the profit tax, the elasticity of reported income remains a sufficient statistic for the optimal tax rate.

Reitering the separation between within market and between market comparisons: the avoidance response is tied to the relationship between firm size and the avoidance cost function. For within market comparative statics, this is directly linked to a market power comparison as well. The across market comparison can only be made if the avoidance costs of the two industry are similar.

4.2 Policy Choice: Taxation vs. Enforcement

The previous section derived the condition for an optimal tax conditional on a not necessarily optimal enforcement rate and vice versa for the optimal enforcement rate. What's equally as important is the optimal balance of these two tools. The condition that governs this tradeoff is

$$\underbrace{(1-\gamma_k) - \frac{\partial \gamma_k}{\partial t_k} t_k + \frac{\alpha}{\lambda}\frac{1}{x_k}\frac{\partial \pi_k}{\partial t_k}}_{\text{Marginal Tax Effect}} = \frac{\rho_k}{\rho^{\beta}_k}\underbrace{\left[-1 - \frac{\partial \gamma_k}{\partial \beta_k} t_k + \frac{\alpha}{\lambda}\frac{1}{x_k}\frac{\partial \pi_k}{\partial \beta_k}\right]}_{\text{Marginal Enforcement Effect}} \tag{27}$$

or, alternatively stated,

$$\underbrace{\left[(1-\gamma_k)+\frac{\rho_k}{\rho_k^{\beta}}\right]}_{\text{Mechanical Change}}+\underbrace{\left[-\frac{\partial\gamma_k}{\partial t_k}+\frac{\partial\gamma_k}{\partial\beta_k}\frac{\rho_k}{\rho_k^{\beta}}\right]t_k}_{\text{Behavioral Change}}+\underbrace{\frac{\alpha}{\lambda}\frac{1}{x_k}\left[\frac{\partial\pi_k}{\partial t_k}-\frac{\partial\pi_k}{\partial\beta_k}\frac{\rho_k}{\rho_k^{\beta}}\right]}_{\text{Profit Change}}=0 \tag{28}$$

While the above expression speaks to the optimum, the LHS of the same expression can be used to dictate whether a policy change consisting of a tax increase and an enforcement decrease is welfare beneficial (if the LHS is greater than 0). In this format, we can easily state the factors that could positively or negatively push for tax rates versus enforcement. Taxes become preferable if (1) the mechanical revenue gain from tax rates is relatively high due to low avoidance rates, (2) the fiscal loss due to avoidance rates increasing both from the increase in taxes and reduction in enforcement is relatively low, and (3) the negative profit responses of taxes are outweighed by the positive profit response from reducing enforcement. All of these factors are impacted by how responsive prices are to enforcement relative to taxes, which dictates how much change in enforcement is needed to compensate for the change in taxes.

The profit effect is the only part of the tradeoff that does not exist in the perfectly competitive version. As profits are zero and remain zero for the competitive industry, this effect is zeroed out. Here, we must consider how raising enforcement versus taxes differentially affects the profits of the industry and feed back to the individual.

We've discussed these factors along the way, but we summarize the main connections here. The more production facilitates avoidance, i.e., the lower the elasticity of marginal cost of avoidance with respect to firm size $\varepsilon^{r}_{H_\gamma}$ is, the more competition drives down avoidance rates. Since avoidance rates are negatively related to the tax base and directly related to the enforcement base, this pushes the tradeoff in favor of taxes at high levels of competition. In other words, if an industry has a low level of avoidance, there is simply more to gain via additional taxation then there is in trying to further reduce avoidance. The relationship between competition and the avoidance response follows in the same direction as discussed in Section 3. While competition may have conflicting impacts on profit margins, ultimately the magnitude of this effect likely does not overpower the previous two effects due to the lower weight on profits since $\frac{\alpha}{\lambda}<1$. This tradeoff can equivalently be demonstrated by combining the two expressions from Proposition 2.

Proposition 3. *The tradeoff between taxes and enforcement follows*

$$\frac{\frac{t_k}{1-t_k}}{\beta_k}=\frac{\left(1-\frac{\alpha}{\lambda}\left[1+EB_k\right]\right)\left(\frac{\alpha}{\lambda}\left[\frac{N_kH_\beta}{Z_k}+EB_k^{\beta}\right]\right)}{t_k\varepsilon_{NR_k}\varepsilon^{\beta}_{NR_k}} \tag{29}$$

where taxable income responses and adminstrative costs are combined into singular net elasticities $\varepsilon^{\theta}_{NR_k}$.

This result is directly comparable to the tradeoff discussed in Keen and Slemrod (2017). Like their conclusion, having a higher value of either elasticity pushes the argument in favor of enforcement, as a higher tax elasticity indicates higher efficiency costs of taxation, while a higher enforcement elasticity indicates more effective enforcement. New to these expressions as compared to their paper are the two excess burden terms. Having a high excess burden for either tool drives down the incentive to use that tool. Recalling the derivation of excess burden, the primary distinction between the two types is in the pass-through rates.

As before, under a pure profit tax, the excess burden terms drop out and only avoidance related behavioral responses remain. For this case then, it can directly be stated that the more firm size facilitates avoidance ($\varepsilon^{r}_{H_\gamma}$ decreases), the more the government should prioritize using enforcement rates over statutory tax rates.

Lastly, regardless of the type of tax, under constant scale avoidance technology, we obtain the unique result that relative desirability of tax versus enforcement is completely independent of the number of firms, and thereby the market power.

Corollary 3.1. *Under constant scale avoidance technology, the optimal ratio of taxation to enforcement for a given industry is independent of the market power in the industry. Under a profit tax, for increasing (decreasing) returns to scale avoidance technology, this ratio decreases (increases) as the market power increases.*

Combining the profit and pass-through expressions from Section 3, the net profit change for a price-neutral policy change under constant scale technology equals zero. This is because the relative erosion of profits via taxes versus via enforcement is exactly equal to relative price effect. Since this is true for any number of firms, this term completely drops out. What this then means is that welfare considerations have zero impact on this tradeoff, and thus the optimization condition is exactly equivalent to if the government had just been maximizing revenue. Second, again using the results from Section 3, the avoidance responses and the pass-through ratio are both independent of firm size. Therefore, the three terms in the breakdown of equation (28) are all independent of firm size and therefore the optimal relationship between taxation and enforcement is independent of market power. These statements are combined into the following corollary.

Corollary 3.2. *Let $\frac{t^*_k}{\beta^*_k}$ represent the ratio of the optimal tax rate to the optimal enforcement rate set by a representative consumer welfare maximizing government in an industry k. Let $\frac{t^R_k}{\beta^R_k}$ represent the ratio of revenue maximizing rates. If the market has firms with constant returns to scale avoidance technology, then*

$$\frac{t^*_k}{\beta^*_k} = \frac{t^R_k}{\beta^R_k}$$

This is true for any exogenous revenue requirement between 0 and the maximum tax revenue possible.

5 Simulations

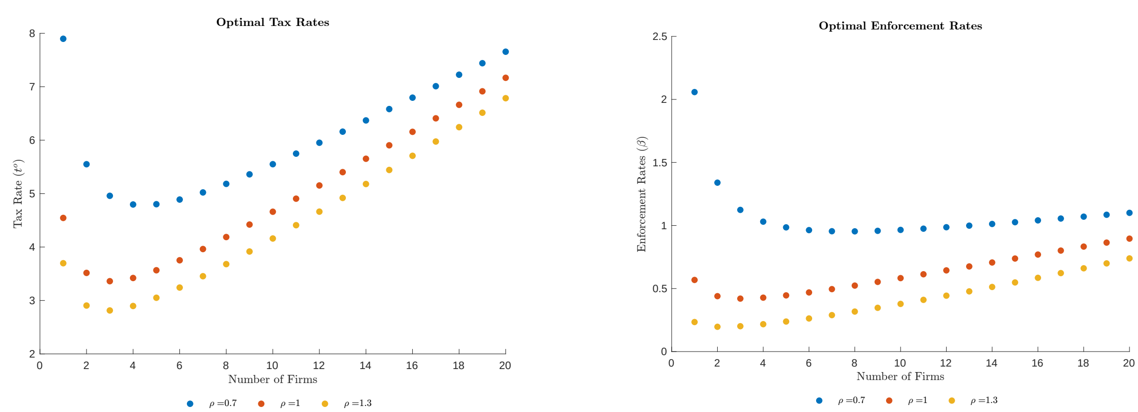
To provide better intution into how the structure of the cost function can affect the results, it is fruitful to illustrate some of these findings with a more concrete cost function. Consider the specifications of the form

$$C(y, r, \gamma, \beta) = \phi(y) + D\beta G(r)\gamma^2$$

where $G(r)$ will be of the form r^ρ. The parameter ρ is then a measure of increasing (cost-reducing) complementarity between the avoidance rate and firm size. In this specification, $\rho = 1$ is equivalent to a elasticity of marginal avoidance costs $\varepsilon^r_{H_\gamma} = 1$, and a higher (lower) ρ increases (decreases) this elasticity. β, in conjunction with the constant D, parameterizes how impactful enforcement can be on the avoidance portion of the firm's cost function. The avoidance fraction γ is squared to allow for strictly convex avoidance costs.

5.1 Simulation Results: Output Tax

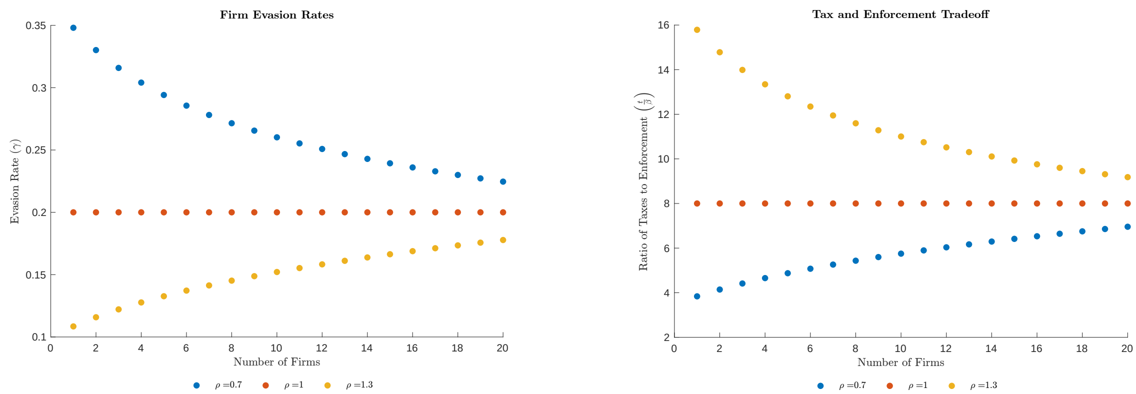
The results of simulations of the model under an output tax are presented here. For the starting simulation environment, I examine variants of a single industry ("within market" comparison) and employ a linear demand curve of the form $q(x) = A - Bx$, and the format of the cost function is as described above. The cost function is further parameterized with squared direct production costs, $F(y) = Ky^2$. The strength of enforcement D is assigned to obtain reasonable values for these illustrations, but ultimately the scale of the y-axes in the following figure are not directly meaningful. Particularly, in the tradeoff figure, a value of "8" implies that taxes are 8 times more valuable than enforcement, but this is highly dependent on the assumed cost of enforcement (normalized to 1) and strength of enforcement parameter.



(a) (b)

Figure 1: Optimal Policy Tools

All simulation results have two layers of heterogeneity. First, each graph should be framed as a comparative static diagram of competition. Number of firms increase along the x-axis, rep-



(a) (b)

Figure 2: Avoidance Rates and Policy Tradeoff

resentative of a reduction in individual firm market power, and we observe how the equilibrium values respond. Second, each figure demonstrates three sets of points to show results at different degrees of firm size-avoidance complementarity. The $\rho = 0.7$ points indicate larger firms having lower avoidance costs, while $\rho = 1.3$ indicates higher costs for larger firms.

Figures 1a and 1b illustrate the transition paths of the optimal tax and enforcement rates. These figures demonstrate that it is possible that tax rates are higher for firms with higher market power in certain regions, while the reverse is true in other regions. The "tipping point" in both figures occurs near the point where ε_{x_k} transitions from a value above one to a value below 1, shifted due to avoidance responses for $\rho \neq 1$, but as previously stated this does not necessarily have to be the case except under stricter conditions.

A key subplot is Figure 2b, which illustrates the tradeoff between taxes and enforcement. The implications of this figure reflects the same conclusions as the previous theory. If size facilitates evasion (ex. $\rho = 0.7$ in the simulations), more market power means the government should shift more priority to enforcement over tax rates. If size reduces the ability to evade (by raising marginal costs faster than benefits), then more market power means the government should have higher priority on tax rates over enforcement.

This figure inversely mirrors Figure 2a on avoidance rates. This aligns with the discussions on the relationship between the tax and enforcement bases as a function of the avoidance rates. When ρ is high, the elasticity of marginal costs of avoidance is high, and increasing competition is tied to higher avoidance rates. High avoidance rates generally reduces the tax rate base and increases the enforcement rate base, which is primarily reflected in the downward trend in the tax-enforcement ratio. The sizes of the potentially counteracting profit effects are not large enough to overturn this primary effect.

6 Uniform Taxation

A government may be limited in its ability to differentiate tax rates. This section thus considers a uniform tax system though still allows enforcement to be targeted. Much of the same intuition still applies, but enforcement must now accomodate previous incentives to differentiate the tax rate. First, the optimal uniform tax rate satisfies

$$\frac{t}{1-t} = \frac{\sum_j Z_j \left[1 - \frac{\alpha}{\lambda}\left[1 + EB_j\right] - \frac{\beta_j}{Z_j}\frac{\partial R_j}{\partial t}\right]}{\sum_j Z_j \varepsilon_{Z_j}} \tag{30}$$

For this uniform tax rate, the correlation between reported income and the elasticity of reported income becomes important. If industries that have high reported incomes are also the ones that have the highest behavioral responses of reported incomes, this leads to a larger loss in the tax base than one would expect if the relationship were random. This is also true for the excess burden term. If the industries with the highest excess burden are the ones that have the highest reported revenue, there is a greater total distortion. Thus, positive correlations between the reported income with either or both the excess burden or the tax elasticity would push the uniform rate lower than it would otherwise be. As before, under a profit tax, this relationship becomes even simpler

$$\frac{t_k}{1-t_k} = \frac{1 - \frac{\alpha}{\lambda}}{\varepsilon_Z} \tag{31}$$

where $\varepsilon_Z = \frac{\sum_j Z_j \varepsilon_{Z_j}}{\sum_j Z_j}$ is the weighted average income elasticity.

Since tax policy cannot be differentiated, all differentiation must be done through the enforcement rates. To get the ratio of enforcement rates across industries, divide the optimal enforcement condition for one market with the same condition for a different market, i.e.,

$$\frac{\beta_m}{\beta_n} = \frac{\frac{H_\beta(z_n,\gamma_n,\beta_n)}{z_n} + EB_n}{\frac{H_\beta(z_m,\gamma_m,\beta_m)}{z_m} + EB_m} \frac{\varepsilon^\beta_{NR_m}}{\varepsilon^\beta_{NR_n}} \tag{32}$$

This is nearly the same as the comparison for a fully differentiated system. The key difference is that at this ratio does not include the ratio in differentiated tax rates by construction. Under the profit tax, the excess burden terms drop out, leaving

$$\frac{\beta_m}{\beta_n} = \frac{\frac{H_\beta(z_n,\gamma_n,\beta_n)}{z_n}}{\frac{H_\beta(z_m,\gamma_m,\beta_m)}{z_m}} \frac{\varepsilon^\beta_{Z_m}}{\varepsilon^\beta_{Z_n}} \tag{33}$$

With differentiated taxes, this expression would be adjusted by the ratio of tax rates $\frac{t_m}{t_n}$. This ratio being greater than 1 implies the government has a reason to target tax rates more heavily in industry m. When taxes are forced to be uniform, the government must instead increase β_m relative to β_n to compensate. If the ratio is below 1, this implies the inverse. The government desires lower relative tax rates in industry m, but cannot do so. Thus, enforcement of industry m

should decrease relative to n. Connecting back to the relationship between firm size and avoidance gives the following proposition:

Proposition 4. *Under a uniform profit tax, if firm size facilitates (hinders) avoidance, $\varepsilon^r_{H_\gamma} < (>) 1$, the enforcement rate is relatively lower (higher) in a more concentrated industry as compared to a less concentrated industry than it would have been in a world with a differentiated profit tax.*

The choice of tax and enforcement for each industry can be equivalently thought of as the choice of effective tax rates and the tax and enforcement ratio. When both tools are perfectly differentiated, the government can exactly select the differentiated effective tax rates that they want in each industry and use the most efficient ratio of tax to enforcement to reach that effective tax rate. When moving to uniform taxes, this is no longer the case. If the government sets effective tax rates targets, they can only reach them by changing the enforcement rate, leading to inefficient ratios. In reverse, if the government only used efficient ratios, they would not be targeting the correct effective tax rates.

Note that with $|K| > 2$, the optimal uniform tax rate may not necessarily "between" the two optimal differentiated tax rates. It could be that in the previous thought experiment, $t > t_m > t_n$. The above proposition speaks to the relative distance between β_m and β_n rather than the absolute difference. It may be possible that on an absolute level, both enforcement rates increase (or decrease) while their relative gap shrinks or expands.

7 Extensions and Alternative Specifications of the Model

This sections consider a few extensions and alternative specifications of the model, particularly in regard to the nature of avoidance and enforcement costs.

7.1 Per-Firm Administrative Costs

In the baseline version of the model, enforcement costs are linear with respect to true industry size. One issue with this assumption is that the government may not be fully aware of true industry size as they may only observe reported income. Thus, this version of the model implicitly assumes that they are able to observe both true income and avoidance rates, but are artifically restricted on only taxing based on true firm size. This may be be a valid assumption if, for example, we believe that all or most of the types of avoidance firms engage in are activities that the government is aware of but may not fully deem worth enforcing.

An alternative version of the enforcement costs that does not assume that the government is aware of avoidance levels is per-firm administrative costs, which only assumes that the government is aware of number of firms in the industry. Thus, the government's budget constraint is

$$\sum_j t_j Z_j - \eta(N_j) a(\beta_j) \tag{34}$$

where $\eta_j(\cdot)$ is a possibly nonlinear, but increasing function of the number of firms, indicative of the rising costs for the government as the number of firms they must potentially enforce increases. In this case, there are straightforward adjustment to the optimum expressions,

$$\frac{t_k}{1-t_k} = \frac{1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right]}{\varepsilon_{Z_k}} \tag{35}$$

and

$$\beta_k = \frac{t_k \varepsilon^{\beta}_{Z_k}}{\frac{\alpha}{\lambda}\left[1 + EB^{\beta}_k\right] + \frac{\eta(N_k)}{Z_k} a_{\beta}} \tag{36}$$

In the tax expression, there is no longer an effect on the tax rate due to the true income changes affecting total enforcement costs. In the enforcement expression, the administrative cost term appears in the denominator, divided by reported income. Total administrative costs increase with the number of firms, but in order for the overall term to increase, it must be the case that these costs increase at a rate faster than the reported income. Thus, as long as $\frac{d}{dN}\left(\frac{\eta}{Z}\right) > 0$, then this provides a downward force on the enforcement rate. If $\eta(N) = N$, then $\frac{\eta}{Z} = \frac{N}{Z} = \frac{1}{(1-\gamma)(yq-\mu F(y))}$. As per firm taxable income decreases with competition, so long as the avoidance benefits of increasing firm size are not too large, this provides an example where the enforcement rate is lower with competition.

On a cross-industry basis, this becomes a simpler comparison. If two industries have the same aggregate reported income, then the enforcement cost factor pushes the enforcement rate down on the one that has a higher number of firms as the average enforcement cost per dollar is higher.

7.2 Advoidance Costs as a Function of Number of Firms

Similar to per-firm administrative costs for the government, number of firms may have an effect on the firm avoidance costs. In the model of the paper, avoidance costs are a function of firm size. An alternative is to directly relate avoidance costs to market power by making avoidance costs a function of number of firms. This type of costs could be framed as an information story. If there are many firms in the same market/industry, then the government has information about what is "normal" in the industry. If there is only a monopoly, the government has no information and must directly enforce this sole firm to learn anything.

Suppose that avoidance costs are given by

$$r_k H(N_k, \gamma_k, \beta_k) \tag{37}$$

where $H_N > 0$ implies that having additional competitors increases the costs of avoidance and the size dependency is assumed linear for simplicity. Then the avoidance FOC is

$$t_k = H_{\gamma}(N_k, \gamma_k, \beta_k) \tag{38}$$

which is independent of firm size but not independent of number of firms. For a given tax policy, increasing the number of firms increases the marginal costs of avoidance. Therefore, it must be the case that the optimally chosen avoidance rate must decrease in order to balance the first order condition.

This is different to having increasing returns to scale avoidance costs in a few ways. First, optimal avoidance rates are now directly tied to market power through number of firms rather than through firm size. Thus, between market comparisons of avoidance rates would directly reflect market power differences. Second, the avoidance response would only have the direct effect. The endogenous size response would drop out since the size of the firm does not impact avoidance rates. The impacts on optimal rates, however, is similar. An industry with higher market power would have higher avoidance rates and be more responsive to both tax policy tools, leading to stronger relative enforcement over taxation.

8 Discussion and Conclusion

Whether market concentration or supernormal firm sizes is cause for additional enforcement attention or higher statutory tax rates has become an increasingly relevant policy question. This paper highlights two primary forces. The first force, the excess burden associated with effective taxation, increases with market power and drives down optimal tax policy rates. The second force, the reported income elasticity, does not have a singular relationship with market power. This leads to an overall ambiguity in the relationship between optimal policy rates and market power. However, there are situations where this relationship can be clearer. For example, for log-concave demand forms production in the inelastic portion of the demand curve implies increasing competition reduces the elasticity of reported income. Thus, optimal rates should increase with competition.

More directly prescriptive is the ratio between optimal taxation and enforcement. The more that market power assists in avoidance, the greater the advantage of enforcement over taxation in concentrated markets. This is because higher avoidance rates imply a relatively low tax base. Therefore, there is more to gain via enforcement and it has a relatively wider range as the more cost-effective tool to raise revenue.

This paper additionally offers some guidance in connecting terms in the optimal expressions to empirics. Fiscal responses are expressed terms of a reported income elasticity, and market related excess burden terms are in terms of pass-through rates, demand elasticities, Lerner indices (isomorphic to number of firms here), price-marginal cost ratios, and marginal complicance costs. Pass-through rates and demand elasticities are common empirical targets. The last two have received relatively less treatment in the past, but have been gaining increased attention in empirical industrial organization research and tax complexity research, respectively.

There are several important extensions to potentially explore. For example, this paper assumes fixed levels of competition in each market. However, it is likely that altering tax policy affects market structure through firm entry or exit, mergers and acquisitions, or collusion. If firms merge in response to stricter enforcement in an attempt to exploit opportunities only available to larger sized firms, this may significantly increase the excess burden of enforcement. Endogenizing market structure responses is important to providing a comprehensive illustration of the role of tax policy in addressing market concentration.

9 References

1. Adachi, Takanori, and Michal Fabinger. 2022. "Pass-through, Welfare, and Incidence Under Imperfect Competition." *Journal of Public Economics*, 211: Article 104589.

2. Allingham, Michael G., and Agnar Sandmo. 1972. "Income Tax Evasion: A Theoretical Analysis." *Journal of Public Economics*, 1: 323-338.

3. Anderson, Simon P., Andrè de Palma, and Brent Kreider. 2001. "Tax Incidence in Differentiated Product Oligopoly." *Journal of Public Economics*, 81: 173-192.

4. Besley, Timothy. "Commodity Taxation and Imperfect Competition." *Journal of Public Economics*, 40: 359-367.

5. Cremer, Helmuth, and Firouz Gahvari. 1990. "Tax Evasion and Optimal Commodity Tax." *Journal of Public Economics*, 50: 261-275

6. de Paula, Áureo, and José A. Scheinkman. 2010. "Value Added Taxes, Chain Effects, and Informality. *American Economic Journal: Macroeconomics*, 2: 195-221.

7. Dharmapala, Dhammika, Joel Slemrod, and John Douglas Wilson. 2011. "Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-Level Administrative Costs." *Journal of Public Economics*, 95: 1036-1047.

8. Eckhout, Jan, Chunyang Fu, Wenjian Li, Xi Weng. 2021. "Optimal Taxation and Market Power." Working Paper.

9. Genakos, Christos, and Mario Pagliero. 2022. "Competition and Pass-Through: Evidence from Isolated Markets." *American Economic Journal: Applied Microeconomics*.

10. Kaplow, Louis. 1990. "Optimal Taxation with Costly Enforcement and Evasion." *Journal of Public Economics*, 43: 221-236.

11. Kaplow, Louis. 2021. "Market Power and Income Taxation." *American Economic Journal: Economic Policy*, 13(4): 329-354.

12. Keen, Michael, and Joel Slemrod. 2017. "Optimal Tax Administration." *Journal of Public Economics*, 152: 133-142.

13. Kopczuk, Wojciech, and Joel Slemrod. 2006. "Putting Firms into Optimal Tax Theory." *The American Economic Review*, 96(2): 130-134.

14. Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod. 2016. "Does Tax Collection Invariance Hold? Evasion and Pass-Through of State Diesel Taxes." *American Economic Journal: Economic Policy*, 8(2): 251-286.

15. Kubick, Thomas R., Daniel P. Lynch, Michael A. Mayberry, and Thomas C. Omer. 2015. "Product Market Power and Tax Avoidance: Market Leaders, Mimicking Strategies, and Stock Returns." *The Accounting Review*, 90(2): 675-702.

16. Martin, Julien, Mathieu Parenti, and Farid Toubal. 2022. "Corporate Tax Avoidance and Sales: Micro Evidence and Aggregate Implications." Working Paper.

17. Miklós-Thal, Jeanine, and Greg Shaffer. 2021. "Pass-Through as and Economic Tool: On Exogenous Competition, Social Incidence, and Price Discriminiation." *Journal of Political Economy*, 129(1): 323-335.

18. Myles, Gareth D. 1989. "Ramsey Tax Rules for Economies with Imperfect Competition." *Journal of Public Economics*, 38:95-115.

19. Pless, Jacquelyn, and Arthur A. van Benthem. 2019. "Pass-Through as a Test for Market Power: An Application for Solar Subsidies." *American Economic Journal: Applied Economics*, 11(4): 367-401.

20. Ramsey, Frank P. 1927. "A Contribution to the Theory of Taxation." *The Economic Journal*, 37(145): 47-61.

21. Ritz, Robert A. "Does Competition Increase Pass-Through?" *The RAND Journal of Economics*, 55(1): 140-165.

22. Slemrod, Joel and Tejaswi Velayudhan. 2018. "Do Firms Remit at Least 85% of Tax Everywhere? New Evidence from India." *Journal of Tax Administration*, 4(1): 24-37.

23. Weyl, E. Glen, and Michal Fabinger. 2013. "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition." *Journal of Political Economy*, 121(3): 528-583.

A Mathematical Appendix

A.1 Firm and Market Responses

In this section, we derive firm and market response margins to each of the tax tools. The relevant response margins are profits, price, and avoidance. Recall that the firm's problem is

$$\max_{y_k,\gamma_k} y_k q_k(x_k) - C_k(y_k, r_k, \gamma_k, \beta_k) - t_k((1-\gamma_k)(y_k q_k - \mu\phi_k(y_k)))$$

To help derive with these responses, recall the first order conditions for production and for avoidance:

$$\begin{aligned}
&y_k : (1-(1-\gamma_k)t_k - H_r)\left[q_k + \frac{\partial q_k}{\partial x_k} y_k\right] = (1-\mu[(1-\gamma_k)t_k + H_r])\phi_y \\
&\gamma_k : t_k(y_k q_k - \mu\phi_k) = H_\gamma
\end{aligned}$$

The first production condition can be rearranged as equating (net) marginal revenues and marginal costs, which will be helpful for later derivations

$$(1-(1-\gamma_k)t_k)\left[q_k + \frac{\partial q_k}{\partial x_k} y_k\right] = (1-\mu(1-\gamma_k)t_k)\phi_y + H_r\left[q_k + \frac{\partial q_k}{\partial x_k} y_k - \mu\phi_y\right]$$

A.1.1 Basic Assumptions on the Cost Function

The Hessian of the profit function $\pi(y, \gamma; t, \beta)$ of a Cournot competitor is

$$H = \begin{bmatrix} -H_{rr}\frac{\partial r_k}{\partial y_k}^2 + (1-\omega_k)\left[N_k\frac{\partial q_k}{\partial x_k} + N_k\frac{\partial^2 q_k}{\partial x_k^2} y_k + \frac{\partial q_k}{\partial x_k}\right] - (1-\mu\omega_k)\phi_{yy} & \frac{\partial r_k}{\partial y_k}(t_k - H_{r\gamma}) \\ \frac{\partial r_k}{\partial y_k}(t_k - H_{r\gamma}) & -H_{\gamma\gamma} \end{bmatrix}$$

where $\omega_k = (1-\gamma_k)t_k + H_r$ and

$$\frac{\partial r_k}{\partial y_k} = q_k + \frac{\partial q_k}{\partial x_k} y_k - \mu\phi_y$$

The determinant of the Hessian is

$$\begin{aligned}
Det(H) \equiv &-\left(\frac{\partial r_k}{\partial y_k}^2\left[-H_{rr}H_{\gamma\gamma} + (t_k - H_{r\gamma})^2\right]\right) \\
&- (1-\omega)\left[N_k\frac{\partial q_k}{\partial x_k} + N_k\frac{\partial^2 q_k}{\partial x_k^2} y_k + \frac{\partial q_k}{\partial x_k}\right] H_{\gamma\gamma} + (1-\mu\omega)\phi_{yy}H_{\gamma\gamma}
\end{aligned}$$

In order for the firm's optimal choice (y_k^*, γ_k^*) to be profit maximizing, it must be true that, evaluated at this choice, $Det(H) > 0$, $\pi_{yy} < 0$ and $\pi_{\gamma\gamma} < 0$, where the first two conditions necessarily imply the third. The third condition requires that $H_{\gamma\gamma} > 0$, which is why the avoidance cost function must be strictly convex in γ.

A.1.2 Changes in Firm Size

Firm size is defined as $r_k = y_k q_k - \mu\phi_k$, i.e., the pre-avoidance production taxable income (simply revenues in output tax case). This is stated trivially in the paper, but I show it more explicitly here. Differentiating the production condition with respect to the number of firms gives

$$\left(-H_{rr}\frac{\partial r_k}{\partial N_k} + (t_k - H_{r\gamma})\frac{\partial \gamma_k}{\partial N_k}\right)\left[q_k + \frac{\partial q_k}{\partial x_k}y_k - \mu\phi_y\right] + (1-\omega)\left[\frac{\partial q_k}{\partial N_k} + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial N_k}y_k + \frac{\partial q_k}{\partial x_k}\frac{\partial y_k}{\partial N_k}\right] - (1-\mu\omega)\phi_{yy}\frac{\partial y_k}{\partial N_k} = 0$$

Differentiating the avoidance condition gives

$$t_k\frac{\partial r_k}{\partial N_k} = H_{\gamma r}\frac{\partial r_k}{\partial N_k} + H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial N_k} \iff \frac{\partial \gamma_k}{\partial N_k}H_{\gamma\gamma} = \frac{\partial r_k}{\partial N_k}(t_k - H_{\gamma r})$$

Plugging this into the previous condition

$$\left(\frac{\partial r_k}{\partial N_k}\left[-H_{rr} + \frac{(t_k - H_{r\gamma})^2}{H_{\gamma\gamma}}\right]\right)\left[q_k + \frac{\partial q_k}{\partial x_k}y_k - \mu\phi_y\right] + (1-\omega)\left[\frac{\partial q_k}{\partial N_k} + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial N_k}y_k + \frac{\partial q_k}{\partial x_k}\frac{\partial y_k}{\partial N_k}\right] - (1-\mu\omega)\phi_{yy}\frac{\partial y_k}{\partial N_k} = 0$$

which can be rearranged as

$$\frac{\partial y_k}{\partial N_k}\left({\frac{\partial r_k}{\partial y_k}}^2\left[-H_{rr} + \frac{(t_k - H_{r\gamma})^2}{H_{\gamma\gamma}}\right]\right) - (1-\mu\omega)\phi_{yy}\frac{\partial y_k}{\partial N_k} + (1-\omega)\left[\frac{\partial q_k}{\partial x_k}\frac{\partial x_k}{\partial N_k} + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial N_k}y_k + \frac{\partial q_k}{\partial x_k}\frac{\partial y_k}{\partial N_k}\right] = 0$$

Using the fact that

$$\frac{\partial x_k}{\partial N_k} = N_k\frac{\partial y_k}{\partial N_k} + y_k$$

we have that

$$\frac{\partial y_k}{\partial N_k}\left[\left({\frac{\partial r_k}{\partial y_k}}^2\left[-H_{rr}H_{\gamma\gamma} + (t_k - H_{r\gamma})^2\right]\right) + (1-\omega)\left[N_k\frac{\partial q_k}{\partial x_k} + N_k\frac{\partial^2 q_k}{\partial x_k^2}y_k + \frac{\partial q_k}{\partial x_k}\right]H_{\gamma\gamma} - (1-\mu\omega)\phi_{yy}H_{\gamma\gamma}\right] + (1-\omega)\frac{\partial q_k}{\partial x_k}y_k\left[1 + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}y_k\right] = 0$$

The term in the brackets is the negative of the determinant of the Hessian of the profit function. Therefore, this expression simplifies to

$$\frac{\partial y_k}{\partial N_k}Det(H) = (1-\omega)\frac{\partial q_k}{\partial x_k}y_k\left[1 + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}y_k\right]$$

Since $Det(H) > 0$ and $\frac{\partial q_k}{\partial x_k} < 0$, in order for the above equality to hold it must be the case that $\frac{\partial y_k}{\partial N_k} < 0$. Thus, we conclude that firm size decreases with competition (or increases with market concentration). As total output decreases with market concentration, and therefore consumer prices increase, it must be the case that per firm revenue and profits also increase.

A.1.3 Proof of Lemma 1

For a multiplicatively separable avoidance cost function $H(r_k, \gamma_k, \beta_k) = G(r_k)I(\gamma_k)J(\beta_k)$, we have that the optimal choice in the avoidance rate satisfies

$$I_\gamma(\gamma_k) = \frac{r_k t_k}{G(r_k)J(\beta)}$$

Since $I_{\gamma\gamma} > 0$, then I_γ is strictly monotone and therefore has an inverse I_γ^{-1}. This implies that

$$\gamma_k^* = I_\gamma^{-1}\left(\frac{r_k t_k}{G(r_k)J(\beta)}\right)$$

where I_γ^{-1} is a strictly increasing function (since we assumed I_γ is). By inspection, this implies that if $G(\cdot)$ is strictly convex, and thus the denominator inside the parentheses increases faster than the numerator, then γ decreases as individual firm size increases. This occurs, in our Cournot framework, when the number of firms increases, i.e., lower market power. Thus, we conclude that a strictly convex $G(\cdot)$ implies decreasing avoidance as market power increases, and a strictly concave $G(\cdot)$ implies increasing avoidance as market power decreases. More formally, since

$$\frac{d}{dN}\frac{r}{G(r)} \propto G(r)\frac{dr}{dN} - rG'\frac{dr}{dN} = G(r)\frac{dr}{dN}\left(1 - \frac{rG'(r)}{G(r)}\right) = G(r)\frac{dr}{dN}\left(1 - \varepsilon^r_{H_\gamma}\right)$$

where $\frac{dr}{dN} < 0$. Then increasing competition reduces optimal avoidance rate iff $\varepsilon^r_{H_\gamma} < 1$. Moreover,

$$\frac{r}{G(r)} = G'(r)\varepsilon^r_{H_\gamma}$$

such that increasing convexity for a given firm size implies a higher value of the above expression. Since optimal avoidance rate is increasing in this expression, increasing convexity implies a higher avoidance rate for a given firm size.

A.1.4 Proof of Lemma 3

Assume $\varepsilon^r_{H_\gamma} < 1$ (all signs are reversed for $\varepsilon^r_{H_\gamma} > 1$). Take the derivative of the direct effect of the avoidance response to taxation with respect to the number of firms:

$$\frac{d}{dN}\frac{r}{(1-\gamma)H_{\gamma\gamma}} = \frac{(1-\gamma)H_\gamma\frac{dr}{dN} - r\left[-\frac{d\gamma}{dN}H_{\gamma\gamma} + (1-\gamma)\frac{dH_{\gamma\gamma}}{dN}\right]}{[(1-\gamma)H_{\gamma\gamma}]^2}$$

Rearranging the numerator (the denominator bears no impact on the sign).

$$\begin{aligned}\frac{d}{dN}\frac{r}{(1-\gamma)H_{\gamma\gamma}} &\propto (1-\gamma)\left(H_{\gamma\gamma}\frac{dr}{dN} - r\frac{dH_{\gamma\gamma}}{dN}\right) + rH_{\gamma\gamma}\frac{d\gamma}{dN} \\ &= (1-\gamma)H_{\gamma\gamma}\frac{dr}{dN}\left[1 - \frac{r}{H_{\gamma\gamma}}\frac{dH_{\gamma\gamma}}{dr}\right] + rH_{\gamma\gamma}\frac{d\gamma}{dN}\end{aligned}$$

where the term in the brackets can use the following application of multiplicative separability of the avoidance cost function:

$$\frac{r}{H_{\gamma\gamma}}\frac{dH_{\gamma\gamma}}{dr} = \frac{r}{G(r)I''(\gamma)J(\beta)}\left[G'(r)I''(\gamma)J(\beta) + G(r)I'''(\gamma)J(\beta)\frac{d\gamma}{dr}\right] = \frac{rG'(r)}{G(r)} + r\frac{I'''}{I''}\frac{d\gamma}{dr}$$

Then

$$\frac{d}{dN}\frac{r}{(1-\gamma)H_{\gamma\gamma}} \propto (1-\gamma)H_{\gamma\gamma}\underbrace{\frac{dr}{dN}}_{<0}\underbrace{\left[1 - \frac{rG'(r)}{G(r)}\right]}_{>0} + \underbrace{rH_{\gamma\gamma}\frac{d\gamma}{dN}\left(1 - (1-\gamma)\frac{I'''}{I''}\right)}_{<0}$$

Under $\varepsilon^{r}_{H_\gamma} < 1$, we have that $\frac{rG'(r)}{G(r)} < 1$. So as long as $(1-\gamma)\frac{I'''}{I''}$ is not too large, this expression is positive. For example $I''' = 0$ when $I(\cdot) = (\cdot)^2$, and is negative for powers below 2. Exponents above 2 are possible, but cannot be strictly guaranteed.

Similarly, taking the derivative of the direct effect of enforcement with respect to the number of firms, we get that

$$\begin{aligned}\frac{d}{dN}\frac{H_{\gamma\beta}}{(1-\gamma)H_{\gamma\gamma}} &\propto (1-\gamma)\left(H_{\gamma\gamma}\frac{dH_{\gamma\beta}}{dN} - H_{\gamma\beta}\frac{dH_{\gamma\gamma}}{dN}\right) + H_{\gamma\beta}H_{\gamma\gamma}\frac{d\gamma}{dN} \\ &= (1-\gamma)H_{\gamma\beta}H_{\gamma\gamma}\underbrace{\frac{dr}{dN}}_{<0}\left(\frac{1}{H_{\gamma\beta}}\frac{dH_{\gamma\beta}}{dr} - \frac{1}{H_{\gamma\gamma}}\frac{dH_{\gamma\gamma}}{dr}\right) + \underbrace{H_{\gamma\beta}H_{\gamma\gamma}\frac{d\gamma}{dN}}_{<0}\end{aligned}$$

where, using our multiplicative separable condition, is weakly positive. Thus, the overall term is negative.

A.1.5 Profit Responses

We differentiate the firm's after-tax profits π_k with respect to the tax rate:

$$\begin{aligned}\frac{\partial \pi_k}{\partial t_k} = (1-H_r)&\left[\frac{\partial y_k}{\partial t_k}q_k + y_k\frac{\partial q_k}{\partial t_k}\right] - \phi_y\frac{\partial y_k}{\partial t_k} + H_r\mu\phi_y\frac{\partial y_k}{\partial t_k} - H_\gamma\frac{\partial \gamma_k}{\partial t_k} - z_k \\ &- t_k\left[-\frac{\partial \gamma_k}{\partial t_k}(y_kq_k - \mu\phi_k) + (1-\gamma_k)\left[\frac{\partial y_k}{\partial t_k}q_k + y_k\frac{\partial q_k}{\partial t_k} - \mu\phi_y\frac{\partial y_k}{\partial t_k}\right]\right]\end{aligned}$$

where as noted in the main body z_k is the firm's reported taxable income. Using the envelope condition, we can simplify this down to

$$\frac{\partial \pi_k}{\partial t_k} = -z_k + (1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\frac{\partial q_k}{\partial t_k}y_k$$

Similarly, a firm's profit change due to an increase in enforcement is given by

$$\frac{\partial \pi_k}{\partial \beta_k} = -H_\beta + (1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\frac{\partial q_k}{\partial \beta_k}y_k$$

The aggregate industry profit (Π_k) responses are then given by

$$\frac{\partial \Pi_k}{\partial t_k} = -Z_k + (1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\frac{\partial q_k}{\partial t_k}x_k$$

$$\frac{\partial \Pi_k}{\partial \beta_k} = -N_k H_\beta + (1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\frac{\partial q_k}{\partial t_k}x_k$$

Under an output tax, we can then use our production FOC to get this in terms of price-cost margins via the Lerner Index. The production FOC can be rearranged as

$$\frac{\phi_y}{q_k} = (1-(1-\gamma_k)t_k - H_r)\left[1 - \frac{1}{N_k \varepsilon_{x_k}}\right]$$

which turns the two profit conditions to

$$\frac{\partial \pi_k}{\partial t_k} = -z_k + \frac{N_k-1}{N_k}\frac{\partial q_k}{\partial t_k}y_k\frac{\phi_y}{q_k}\frac{1}{1-\frac{1}{N_k\varepsilon_{x_k}}}$$

$$\frac{\partial \pi_k}{\partial \beta_k} = -H_\beta + \frac{N_k-1}{N_k}\frac{\partial q_k}{\partial \beta_k}y_k\frac{\phi_y}{q_k}\frac{1}{1-\frac{1}{N_k\varepsilon_{x_k}}}$$

A.1.6 Price Responses

To obtain the price responses, we must differentiate the first order condition of production with respect to each of the two tax tools. As before, we start with the tax rate:

$$\begin{aligned}
&\left(\frac{\partial \gamma_k}{\partial t_k}t_k - (1-\gamma_k) - \left[H_{rr}\left[\frac{\partial y_k}{\partial t_k}q_k + y_k\frac{\partial q_k}{\partial t_k} - \mu\phi_y\frac{\partial y_k}{\partial t_k}\right] + H_{r\gamma}\frac{\partial \gamma_k}{\partial t_k}\right]\right)\left(q_k + \frac{\partial q_k}{\partial x_k}y_k\right) \\
&\qquad\qquad\qquad +(1-\omega_k)\left(\frac{\partial q_k}{\partial t_k} + \frac{\partial y_k}{\partial t_k}\frac{\partial q_k}{\partial x_k} + y_k\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial t_k}\right) \\
&= \mu\left(\frac{\partial \gamma_k}{\partial t_k}t_k - (1-\gamma_k) - \left[H_{rr}\left[\frac{\partial y_k}{\partial t_k}q_k + y_k\frac{\partial q_k}{\partial t_k} - \mu\phi_y\frac{\partial y_k}{\partial t_k}\right] + H_{r\gamma}\frac{\partial \gamma_k}{\partial t_k}\right]\right)\phi_y \\
&\qquad\qquad\qquad +(1-\mu\omega_k)\phi_{yy}\frac{\partial y_k}{\partial t_k}
\end{aligned}$$

which we can rearrange as

$$\frac{\partial q_k}{\partial t_k}\left[(1-\omega_k)\left(1+\frac{1}{N_k}\left[1+\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}\right]\right)-(1-\mu\omega_k)\frac{\phi_{yy}}{N_k}\frac{\partial x_k}{\partial q_k}-\frac{H_{rr}}{N_k}\frac{\partial x_k}{\partial q_k}\left(q_k+\frac{\partial q_k}{\partial x_k}y_k-\mu\phi_y\right)^2\right]$$
$$=\left(1-\gamma_k-\frac{\partial \gamma_k}{\partial t_k}(1-\mu\phi_y)\left[t_k-H_{r\gamma}\right]\right)\left(q_k+\frac{\partial q_k}{\partial x_k}y_k-\mu\phi_y\right)$$

Before moving on, first note that if $\mu=1$, then the original expression simplifies to

$$\left(\frac{\partial \gamma_k}{\partial t_k}t_k-(1-\gamma_k)-\left[H_{rr}\left[\frac{\partial y_k}{\partial t_k}q_k+y_k\frac{\partial q_k}{\partial t_k}-\mu\phi_y\frac{\partial y_k}{\partial t_k}\right]+H_{r\gamma}\frac{\partial \gamma_k}{\partial t_k}\right]\right)\left(q_k+\frac{\partial q_k}{\partial x_k}y_k-\phi_y\right)$$
$$(1-\omega_k)\left(\frac{\partial q_k}{\partial t_k}+\frac{\partial y_k}{\partial t_k}\frac{\partial q_k}{\partial x_k}+y_k\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial t_k}-\phi_{yy}\frac{\partial y_k}{\partial t_k}\right)=0$$

The first line is zeroed since the second bracketed term is zero by the production condition. Then, this leaves

$$(1-\omega_k)\frac{\partial q_k}{\partial t_k}\left(1+\frac{1}{N_k}\left(1+\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial t_k}x_k-\phi_{yy}\frac{\partial x_k}{\partial q_k}\right)\right)=0$$

which implies that $\frac{\partial q_k}{\partial t_k}=0$. This makes sense as the output decision is independent of the tax rate and avoidance decision when costs are fully deductible. The same conclusion can be found for the pass-through rate of the enforcement tool.

Returning to the case where $\mu<1$, new elasticities will be introduced in order to simplify the expression. First, as in Weyl and Fabinger (2013), define the (negative) marginal surplus of quantity expansion as $ms_k=\frac{\partial q_k}{\partial x_k}y_k=\frac{x_k}{N_k}\frac{\partial q_k}{\partial x_k}$. Then

$$\frac{\partial ms_k}{\partial x_k}=-\frac{1}{N_k}\frac{\partial q_k}{\partial x_k}\left[1+\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}x_k\right]$$

Then the inverse of the elasticity of marginal surplus is

$$\begin{aligned}\frac{1}{\varepsilon_{ms_k}}&=\frac{\partial ms_k}{\partial x_k}\frac{x_k}{ms_k}\\&=-\frac{1}{N_k}\frac{\partial q_k}{\partial x_k}\left[1+\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}x_k\right]\frac{N_k}{\frac{\partial q_k}{\partial x_k}}\\&=1+\frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial q_k}x_k\end{aligned}$$

Second, the net marginal costs as

$$\begin{aligned}MC_k&=(1-\mu(1-\gamma_k)t_k)\phi_y+H_r\left[q_k+\frac{\partial q_k}{\partial x_k}y_k-\mu\phi_y\right]\\&=(1-\mu(1-\gamma_k)t_k)\phi_y+H_r q_k\left[1-\frac{1}{N_k\varepsilon_{x_k}}-\mu\frac{\phi_y}{q_k}\right]\end{aligned}$$

Using the production FOC, this gives

$$\frac{MC_k}{q_k} = (1 - (1 - \gamma_k)t_k) \left[1 - \frac{1}{N_k \varepsilon_{x_k}}\right]$$

We then define the elasticty of the inverse marginal cost curve (a pseudo "supply") as

$$\begin{aligned}
\frac{1}{\varepsilon_{s_k}} \frac{MC_k}{x_k} &= \frac{\partial MC_k}{\partial x_k} \\
&= (1 - \mu(1 - \gamma_k)t_k) \frac{1}{N_k} \phi_{yy} + \frac{1}{N_k} H_{rr} \left[q_k + x_k \frac{\partial q_k}{\partial x_k} - \mu \phi_y\right]^2 \\
&\quad + H_r \left[\frac{\partial q_k}{\partial x_k} + \frac{1}{N_k} \left(\frac{\partial q_k}{\partial x_k} + x_k \frac{\partial^2 q_k}{\partial x_k^2}\right) - \mu \frac{1}{N_k} \phi_{yy}\right] \\
&= ((1 - \mu[(1 - \gamma_k)t_k + H_r)]) \frac{1}{N_k} \phi_{yy} + \frac{1}{N_k} H_{rr} \left[q_k + x_k \frac{\partial q_k}{\partial x_k} - \mu \phi_y\right]^2 \\
&\quad + H_r \left[\frac{\partial q_k}{\partial x_k} + \frac{1}{N_k} \left(\frac{\partial q_k}{\partial x_k} + x_k \frac{\partial^2 q_k}{\partial x_k^2}\right)\right]
\end{aligned}$$

Multiplying through by $\frac{\partial x_k}{\partial q_k}$, then we can convert our pass-through expression as

$$\begin{aligned}
\frac{\partial q_k}{\partial t_k} \left[(1 - (1 - \gamma_k)t_k) \left(1 + \frac{1}{N_k} \left[1 + \frac{\partial^2 q_k}{\partial x_k^2} \frac{\partial x_k}{\partial q_k}\right]\right) + \frac{\varepsilon_{x_k}}{\varepsilon_{s_k}} \frac{MC_k}{q_k}\right] \\
= \left(1 - \gamma_k - \frac{\partial \gamma_k}{\partial t_k} [t_k - H_{r\gamma}]\right) \left(q_k + \frac{\partial q_k}{\partial x_k} y_k\right)
\end{aligned}$$

and plugging in our value of $\frac{MC_k}{q_k}$ and the marginal surplus elasticity from before,

$$\begin{aligned}
\frac{\partial q_k}{\partial t_k} \left[(1 - (1 - \gamma_k)t_k) \left(1 + \frac{1}{N_k \varepsilon_{ms_k}} + \frac{\varepsilon_{x_k} - \frac{1}{N_k}}{\varepsilon_{s_k}}\right)\right] \\
= \left(1 - \gamma_k - \frac{\partial \gamma_k}{\partial t_k} (1 - \mu \phi_y) [t_k - H_{r\gamma}]\right) \left(q_k + \frac{\partial q_k}{\partial x_k} y_k\right)
\end{aligned}$$

The process for the pass-through rate of the enforcement parameter is very similar. Thus, we can write our final pass-through rates as

$$\frac{\partial q_k}{\partial t_k} = \frac{\left(1 - \gamma_k - \frac{\partial \gamma_k}{\partial t_k} [t_k - H_{r\gamma}]\right) \left(q_k + \frac{\partial q_k}{\partial x_k} y_k - \mu \phi_y\right)}{(1 - (1 - \gamma_k)t_k) \left(1 + \frac{1}{N_k \varepsilon_{ms_k}} + \frac{\varepsilon_{x_k} - \frac{1}{N_k}}{\varepsilon_{s_k}}\right)}$$

$$\frac{\partial q_k}{\partial \beta_k} = \frac{\left(H_{\gamma\beta} - \frac{\partial \gamma_k}{\partial \beta_k} [t_k - H_{r\gamma}]\right) \left(q_k + \frac{\partial q_k}{\partial x_k} y_k - \mu \phi_y\right)}{(1 - (1 - \gamma_k)t_k) \left(1 + \frac{1}{N_k \varepsilon_{ms_k}} + \frac{\varepsilon_{x_k} - \frac{1}{N_k}}{\varepsilon_{s_k}}\right)}$$

A.1.7 Avoidance Response

For the avoidance response, differentiating the avoidance first order condition with respect to the tax rate gives

$$r_k + (t_k - H_{\gamma r})\frac{\partial r_k}{\partial t_k} = H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial t_k}$$

Expanding the change in firm size,

$$r_k + (t_k - H_{\gamma r})\left(\frac{\partial y_k}{\partial t_k}q_k + y_k\frac{\partial q_k}{\partial t_k} - \mu\phi_y\frac{\partial y_k}{\partial t_k}\right) = H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial t_k}$$

which can be rearranged as

$$H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial t_k} = r_k + (t_k - H_{\gamma r})y_k\frac{\partial q_k}{\partial t_k}\left(1 - \varepsilon_{x_k} + \mu\frac{\phi_y}{q_k}\varepsilon_{x_k}\right)$$

or

$$\frac{\partial \gamma_k}{\partial t_k} = \frac{r_k}{H_{\gamma\gamma}} + \frac{(t_k - H_{\gamma r})}{H_{\gamma\gamma}}y_k\frac{\partial q_k}{\partial t_k}\left[1 - \left(1 - \mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]$$

The corresponding enforcement condition is

$$(t_k - H_{\gamma r})\frac{\partial r_k}{\partial t_k} = H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial \beta_k} + H_{\gamma\beta}$$

which can be rearranged as

$$\frac{\partial \gamma_k}{\partial \beta_k} = -\frac{H_{\gamma\beta}}{H_{\gamma\gamma}} + \frac{(t_k - H_{\gamma r})}{H_{\gamma\gamma}}y_k\frac{\partial q_k}{\partial \beta_k}\left[1 - \left(1 - \mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]$$

Lastly, reconsider the avoidance FOC

$$t_k r_k = H_\gamma$$

Using the definition of the elasticity of marginal avoidance costs, this implies that

$$t_k - H_{\gamma r} = t_k\left(1 - \frac{H_{\gamma r}}{t_k}\right) = t_k\left(1 - H_{\gamma r}\frac{H_\gamma}{r}\right) = t_k(1 - \varepsilon^r_{H_\gamma})$$

Thus,

$$t_k - H_{\gamma r} > 0 \iff \varepsilon^r_{H_\gamma} < 1$$

A.1.8 Welfare Effect

The effect of a change in one of the tax tools on welfare is the effect on the representative consumer, i.e.,

$$\frac{dW}{d\theta_k} = \alpha\left[-\frac{\partial q_k}{\partial \theta_k}x_k + \frac{\partial \Pi_k}{\partial \theta_k}\right]$$

where $\theta_k \in \{t_k, \beta_k\}$. Starting with the tax rate, plugging in the value for the profit change calculated previously gives

$$\frac{1}{\alpha Z_k}\frac{dW}{dt_k} = -\left[1 + \frac{\partial q_k}{\partial t_k}\frac{x_k}{Z_k}\left[1 - (1-(1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\right]\right]$$

Define the "1" as the mechanical dollar raised, while the rest of the expression is the excess burden. Thus, this expression simplifies to

$$\frac{1}{\alpha Z_k}\frac{dW}{dt_k} = -\left[1 + EB_k\right]$$

Similarly, for the enforcement, the welfare effect is

$$\frac{1}{\alpha Z_k}\frac{dW}{d\beta_k} = -\left[\frac{N_k H_\beta}{Z_k} + \frac{\partial q_k}{\partial \beta_k}\frac{x_k}{Z_k}\left[1 - (1-(1-\gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\right]\right]$$

where an excess burden type term can be equivalently defined, and the previous expression rewritten as

$$\frac{1}{\alpha Z_k}\frac{dW}{d\beta_k} = -\left[\frac{N_k H_\beta}{Z_k} + EB_k^\beta\right]$$

Under a pure profit tax where $\mu = 1$, the price pass-through is zero for both tools, and therefore, $EB_k = EB_k^\beta = 0$. Thus, the per taxable income welfare effect of taxation is independent of competition (and firm size). Meanwhile, the welfare effect of enforcement may still have a dependency on market power as $\frac{H_\beta}{z_k}$ is dependent on the avoidance rate. This relationship is discussed more in Lemma 2.1.

A.1.9 Taxable Income Effect

The relevant elasticity is the elasticity of taxable income, which is reported revenue for and output tax ($\mu = 0$) case and reported profits in the profit tax case ($\mu = 1$). Then the defined elasticity of reported taxable income in the paper can be derived as

$$\begin{aligned}\frac{\varepsilon_{Z_k}}{1-t_k} &= \frac{1}{Z_k}\frac{\partial Z_k}{\partial(1-t_k)} \\ &= \frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{R_k}{Z_k} - \frac{(1-\gamma_k)x_k}{Z_k}\frac{\partial q_k}{\partial t_k}\left[1-\varepsilon_{x_k}+\mu\frac{\phi_y}{q_k}\varepsilon_{x_k}\right] \\ &= \frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{1}{1-\gamma_k} - \frac{(1-\gamma_k)x_k}{Z_k}\frac{\partial q_k}{\partial t_k}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]\end{aligned}$$

Under a profit tax, the second term drops out, leaving

$$\varepsilon_{Z_k} = \varepsilon_{1-\gamma_k}$$

i.e., that the elasticity of reported taxable income is exactly equal to the tax elasticity of avoidance rate. In the main body of the paper the avoidance elasticity is labeled as ε_{γ_k} instead of $\varepsilon_{1-\gamma_k}$ for shorthand though the latter is more explicit. Similarly, the elasticity of enforcement is defined by

$$\begin{aligned}\frac{\varepsilon_{Z_k}^{\beta_k}}{\beta_k} &= \frac{1}{Z_k}\frac{\partial Z_k}{\partial\beta_k} \\ &= \frac{\partial(1-\gamma_k)}{\partial\beta_k}\frac{R_k}{Z_k} + \frac{(1-\gamma_k)x_k}{Z_k}\frac{\partial q_k}{\partial\beta_k}\left[1-\varepsilon_{x_k}-\mu\frac{\phi_y}{q_k}\varepsilon_{x_k}\right] \\ &= \frac{\partial(1-\gamma_k)}{\partial\beta_k}\frac{1}{1-\gamma_k} + \frac{(1-\gamma_k)x_k}{Z_k}\frac{\partial q_k}{\partial\beta_k}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]\end{aligned}$$

and once again, under a profit tax the enforcement elasticity of taxable income is equal to enforcement elasticity of avoidance

$$\varepsilon_{Z_K}^{\beta_k} = \varepsilon_{1-\gamma_k}^{\beta_k}$$

Returning to the tax expression, plugging in the expression for the avoidance response gives

$$\begin{aligned}\frac{\varepsilon_{Z_k}}{1-t_k} &= \left(\frac{r_k}{H_{\gamma\gamma}} + \frac{(t_k-H_{\gamma r})}{H_{\gamma\gamma}}y_k\frac{\partial q_k}{\partial t_k}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]\right)\frac{1}{1-\gamma_k} \\ &\quad - \frac{(1-\gamma_k)x_k}{Z_k}\frac{\partial q_k}{\partial t_k}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right] \\ &= \frac{r_k}{(1-\gamma_k)H_{\gamma\gamma}} + \left(\frac{(t_k-H_{\gamma r})y_k}{(1-\gamma_k)H_{\gamma\gamma}} - \frac{(1-\gamma_k)x_k}{Z_k}\right)\frac{\partial q_k}{\partial t_k}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]\end{aligned}$$

Under a pure profit tax, all size based changes drop out of the expression leaving only the direct effect on the avoidance response. The same is true for the enforcement elasticity.

A.1.10 Proposition 1

Consider an output tax as there is no change in reported income under a profit tax. The exercise is to differentiate the elastity of reported revenue with respect to the number of firms. Under constant avoidance evasion technology, the first term is independent of firm size, and therefore independent of the number of firms. Thus, the relationship hinges on how the second term changes with a rising number of firms. More directly, the relevant expression for the tax elasticity is

$$\frac{d}{dN_k}\left(\frac{\partial q_k}{\partial t_k^o}\frac{1}{q_k}(\varepsilon_{x_k}-1)\right)=\frac{d}{dN_k}\left(\frac{\partial q_k}{\partial t_k^o}\frac{1}{q_k}\right)(\varepsilon_{x_k}-1)+\frac{\partial q_k}{\partial t_k^o}\frac{1}{q_k}\frac{d}{dN_k}(\varepsilon_{x_k}-1)$$

where a positive value of the above expression implies a larger elasticity. Under log-concave demand, the pass-through rate increases with the number of firms, meaning that the first term is negative when $\varepsilon_{x_k}<1$. Similarly, the demand elasticity is decreasing with the number of firms, which means the second term is also negative. Thus, the overall expression is negative, meaning the elasticity of reported revenue is smaller.

Similarly, differentiate the elastity of reported revenue of enforcement with respect to the number of firms. Under constant avoidance evasion technology, this expression similarly solely relies on the behavior of the true income response. The difference positively, such that the relevant expression is

$$\frac{d}{dN_k}\left(\frac{\partial q_k}{\partial t_k^o}\frac{1}{q_k}(1-\varepsilon_{x_k})\right)=\frac{d}{dN_k}\left(\frac{\partial q_k}{\partial \beta_k}\frac{1}{q_k}\right)(1-\varepsilon_{x_k})+\frac{\partial q_k}{\partial \beta_k}\frac{1}{q_k}\frac{d}{dN_k}(1-\varepsilon_{x_k})$$

Under log-concave demand, the pass-through rate increases with the number of firms, meaning that the first term is positive when $\varepsilon_{x_k}<1$. Similarly, the demand elasticity is decreasing with the number of firms, which means the second term is also positive.

A.1.11 A Simple Example for Reported Income Elasticity

To illustrate the previous result, first ignore the avoidance portion of the issue and assume a demand function with zero log concavity and marginal costs are constant. Then under an output tax, the pass-through rate is

$$\rho_k=\frac{1-\frac{1}{N_k\varepsilon_{x_k}}}{(1-t_k)\left(1+\frac{1}{N_k}\right)}$$

Then

$$
\begin{aligned}
(1-t_k)\frac{d}{dN_k}\rho_k &= \frac{\left(1+\frac{1}{N}\right)\left(\frac{\varepsilon_{x_k}+N\frac{d\varepsilon_{x_k}}{dN}}{(N_k\varepsilon_{x_k})^2}\right)-\left(1-\frac{1}{N_k\varepsilon_{x_k}}\right)\left(-\frac{1}{N_k^2}\right)}{\left(1+\frac{1}{N_k}\right)^2} \\
&= \frac{\left(\frac{\varepsilon_{x_k}+N\frac{d\varepsilon_{x_k}}{dN}}{(N_k\varepsilon_{x_k})^2}\right)}{\left(1+\frac{1}{N_k}\right)}+\rho_k\frac{1}{N_k^2+N_k} \\
&= \frac{\left(\frac{\varepsilon_{x_k}+N\frac{d\varepsilon_{x_k}}{dN}}{(N_k\varepsilon_{x_k})^2}\right)}{\left(1+\frac{1}{N_k}\right)}+\rho_k\frac{1}{N_k^2+N_k}
\end{aligned}
$$

which means that

$$
(1-t_k)\left(\frac{d}{dN_k}\rho_k\right)(\varepsilon_{x_k}-1)=\frac{\frac{1}{N^2}\left(1-\frac{1}{\varepsilon_{x_k}}\right)+\frac{1}{N_k\varepsilon_{x_k}}\left(1-\frac{1}{\varepsilon_{x_k}}\right)\frac{d\varepsilon_{x_k}}{dN_k}}{\left(1+\frac{1}{N_k}\right)}+\rho_k\frac{\varepsilon_{x_k}-1}{N_k^2+N_k}
$$

On the other hand,

$$
\rho_k\frac{d}{dN_k}(\varepsilon_{x_k}-1)=\rho_k\frac{d\varepsilon_{x_k}}{dN_k}
$$

Start at a monopoly such that $N_k=1$. Then

$$
(1-t_k)\left(\frac{d}{dN_k}\rho_k\right)(\varepsilon_{x_k}-1)=\rho_k+\frac{1}{\varepsilon_{x_k}}\rho_k\frac{d\varepsilon_{x_k}}{dN_k}
$$

A.2 Perfectly Competitive Benchmarks

In this section, we benchmark market responses to the competitive limit. First, the production FOC reduces to

$$
(1-(1-\gamma_k)t_k-H_r)q_k=(1-\mu[(1-\gamma_k)t_k+H_r])\phi_y
$$

while the avoidance condition has the same form as before.

A.2.1 Price Response

To get the price response for a competitive market, rather than differentiating the production condition, we utilize the zero profit condition. For a change in the tax tool, the change in profits must remain zero, so we have by the envelope condition that

$$
y_k\frac{\partial q_k}{\partial t_k}-H_r\left[y_k\frac{\partial q_k}{\partial t_k}\right]-z_k-t_k\left[(1-\gamma_k)\left[y_k\frac{\partial q_k}{\partial t_k}\right]\right]=0
$$

which we can rearrange as

$$\frac{\partial q_k}{\partial t_k} = \frac{z_k}{y_k(1 - t_k(1-\gamma_k) - H_r)}$$

The enforcement pass-through in a competitive market is

$$\frac{\partial q_k}{\partial \beta_k} = \frac{H_\beta}{y_k(1 - t_k(1-\gamma_k) - H_r)}$$

Under a pure profit tax $\gamma_k^* = 0$, and $z_k = r_k = 0$. Therefore, the tax pass-through is zero (since $z_k = 0$), and the enforcement pass-through is also zero (since $\gamma_k^* = 0$).

A.2.2 Welfare

Since the change in profits is zero, the only effect is the change in the consumer's income due to the price effect. This means

$$\frac{1}{\alpha Z_k}\frac{dv}{dt_k} = -\frac{\partial q_k}{\partial t_k}\frac{x_k}{Z_k} = -1 - \left[\frac{\partial q_k}{\partial t_k}\frac{x_k}{Z_k} - 1\right]$$

$$\frac{1}{\alpha Z_k}\frac{dv}{d\beta_k} = -\frac{\partial q_k}{\partial \beta_k}\frac{x_k}{Z_k} = -\frac{N_k H_\beta}{Z_k} - \left[\frac{\partial q_k}{\partial \beta_k}\frac{x_k}{Z_k} - \frac{N_k H_\beta}{Z_k}\right]$$

Plugging in the pass-through expressions we found above, we get

$$\frac{1}{\alpha Z_k}\frac{dv}{dt_k} = -\frac{1}{1 - t_k(1-\gamma_k) - H_r} = -1 - \left[\frac{t_k(1-\gamma_k) + H_r}{1 - t_k(1-\gamma_k) - H_r}\right]$$

$$\frac{1}{\alpha Z_k}\frac{dv}{d\beta_k} = -\frac{N_k H_\beta / Z_k}{1 - t_k(1-\gamma_k) - H_r} = -\frac{N_k H_\beta}{Z_k} - \left[\frac{\frac{N_k H_\beta}{Z_k}(t_k(1-\gamma_k) + H_r)}{1 - t_k(1-\gamma_k) - H_r}\right]$$

In both expressions, the term in the bracket is what replaces the excess burden factor in the more general expressions.

A.2.3 Taxable Income

The outward form of the change in reported taxable income is identical to the general expression.

A.3 Core Results

In this section, we derive the core results of the paper–primarily the optimal tax and enforcement level and tradeoff expressions.

A.3.1 Optimal Levels (Proposition 2)

We begin with the government's Lagrangian:

$$\mathcal{L} = v(q) + \lambda \left[\sum_j t_j Z_j - \beta_j R_j - G \right]$$

Differentiating this expression with respect to the tax rate of an inudstry k, we get

$$\alpha \left[-\frac{\partial q_k}{\partial t_k} x_k + \frac{\partial \Pi_k}{\partial t_k} \right] + \lambda \left[Z_k + t_k \frac{\partial Z_k}{\partial t_k} - \beta_k \frac{\partial R_k}{\partial t_k} \right]$$

Dividing through the previous expression by reported income, we get

$$\frac{\alpha}{\lambda} \frac{1}{Z_k} \left[-\frac{\partial q_k}{\partial t_k} x_k + \frac{\partial \Pi_k}{\partial t_k} \right] + 1 - \frac{t_k}{1 - t_k} \varepsilon_{Z_k} - \beta_k \frac{1}{z_k} \frac{\partial R_k}{\partial t_k}$$

where the elasticity of reported taxable income is defined with respect to the retention rate $1 - t_k$. Using our work from previous section, we define

$$1 + EB_k = \frac{1}{z_k} \left[\frac{\partial q_k}{\partial t_k} x_k - \frac{\partial \Pi_k}{\partial t_k} \right]$$

and thus, write this tax result in the simple form

$$\frac{t_k}{1 - t_k} = \frac{1 - \frac{\alpha}{\lambda} \left[1 + EB_k \right] - \frac{\beta_k}{z_k} \frac{\partial R_k}{\partial t_k}}{\varepsilon_{z_k}}$$

or, if we include enforcement costs with the taxable income elasticity to get a "net revenue elasticity", we can write this as

$$\frac{t_k}{1 - t_k} = \frac{1 - \frac{\alpha}{\lambda} \left[1 + EB_k \right]}{\varepsilon_{NR_k}}$$

Similarly, on the enforcement side, we can differentiate the expression with respect to enfocement to get

$$\alpha \left[-\frac{\partial q_k}{\partial \beta_k} x_k + \frac{\partial \Pi_k}{\partial \beta_k} \right] + \lambda \left[t_k \frac{\partial Z_k}{\partial \beta_k} - R_k - \beta_k \frac{\partial R_k}{\partial \beta_k} \right]$$

which we can similarly convert to

$$-\frac{\alpha}{\lambda} \left[\frac{N_k H_\beta}{Z_k} + EB_k^\beta \right] + \frac{t_k}{\beta_k} \varepsilon_{Z_k}^{\beta_k} - \frac{R_k}{Z_k} - \frac{\beta_k}{Z_k} \frac{\partial R_k}{\partial \beta_k}$$

giving us

$$\beta_k = \frac{t_k \varepsilon_{z_k}^{\beta_k}}{\frac{R_k}{z_k} + \frac{\beta_k}{z_k}\frac{\partial R_k}{\partial \beta_k} + \frac{\alpha}{\lambda}\left[\frac{N_k H_\beta}{Z_k} + EB_k^\beta\right]}$$

or, once again using a net revenue elasticity that includes enforcement costs into the income elasticity,

$$\beta_k = \frac{t_k \varepsilon_{NR_k}^{\beta_k}}{\frac{\alpha}{\lambda}\left[\frac{H_\beta}{z_k} + EB_k^\beta\right]}$$

Alternatively, instead of explicity finding an enforcement rate, we simply illustrate that the enforcement elasticity is

$$\varepsilon_{z_k}^{\beta} = \frac{\beta_k \frac{\alpha}{\lambda}\left[N_k H_\beta + Z_k EB_k^\beta\right] + \beta_k R_k + \beta_k^2 \frac{\partial R_k}{\partial \beta_k}}{t_k Z_k}$$

as in Keen and Slemrod (2017). In this form, the RHS is, as they describe, an adjusted cost-to-revenue ratio and thus the elasticity of enforcement fully captures the tradeoff. The costs are a sum of the welfare cost, compliance costs for the firm and excess burden on the market, and the total administrative cost.

A.3.2 Corollary 2.1

Under a profit tax ($\mu = 1$), we have that $EB_k = EB_K^\beta = 0$. Additionally, we have that $\varepsilon_{Z_k} = \varepsilon_{1-\gamma_k}$ and $\varepsilon_{Z_k}^\beta = \varepsilon_{1-\gamma_k}^\beta$. Thus, we can simplify our two prior expressions down to

$$\frac{t_k}{1-t_k} = \frac{1-\frac{\alpha}{\lambda}}{\varepsilon_{\gamma_k}}, \qquad \beta_k = \frac{t_k \varepsilon_{\gamma_k}^{\beta_k}}{\frac{R_k}{Z_k} + \frac{\alpha}{\lambda}\frac{H_\beta}{z_k}}$$

The statements on the relationship between avoidance technology, market power, and the optimal tax rates directly follow Lemma 3. Under constant scale avoidance, we have that ε_{γ_k} is independent of the market power within a market. Thus, the tax rate does not depend on market concentration. For increasing returns to scale avoidance technology, the behavioral response is larger for larger firms, and thus the tax rate is driven down as market power increases.

Similarly, $\varepsilon_{\gamma_k}^\beta$ is independent of market power under constant scale avoidance and related to the complementarity between firm size and avoidance costs: increasing returns to scale avoidance technology implies a higher avoidance elasticity. Under constant scale avoidance, we have that $\frac{H_\beta}{z_k} \not\propto z_k$, and therefore the enforcement level is also independent of market power. The relationship with optimal enforcement rates in other cases is more nuanced due to the effect on profits being in a different direction than the effect on the avoidance response. Consider the change in welfare inclusive of revenue

$$\begin{aligned}\frac{dW}{d\beta} &= -\frac{\alpha}{\lambda}\frac{H_\beta}{z_k} + \frac{t_k}{\beta_k}\varepsilon^{\beta}_{\gamma_k} - \frac{R_k}{Z_k} = 0 \\ &= -\frac{\alpha}{\lambda}\frac{H_\beta}{r_k} + \frac{t_k(1-\gamma_k)}{\beta_k}\varepsilon^{\beta}_{\gamma_k} - 1 = 0 \\ &= -\frac{\alpha}{\lambda}\frac{H_\beta}{r_k} + t_k\frac{H_{\gamma\beta}}{H_{\gamma\gamma}} - 1 = 0\end{aligned}$$

Since

$$\frac{H_\beta}{r}\frac{H_\gamma}{H_\gamma} = t\frac{H_\beta}{H_\gamma}$$

The last line becomes

$$t_k\left[\frac{H_{\gamma\beta}}{H_{\gamma\gamma}} - \frac{\alpha}{\lambda}\frac{H_\beta}{H_\gamma}\right] - 1 = 0$$

Assume that $\varepsilon_{H^r_\gamma} < 1$. If N_k increases, the middle brackets becomes smaller, and β must decrease to keep the condition equal to zero. Thus, the optimal β decreases with respect to competition under this assumption on the cost function.

A.3.3 Tradeoff (Proposition 3)

We will now also get our expression for the optimal tradeoff, which we will do so in two ways. First, we will simply divide our previous two expressions and obtain a Keen and Slemrod (2017) style tradeoff between policy elasticities in addition to the excess burden costs.

$$\frac{\frac{t_k}{1-t_k}}{\beta_k} = \frac{\left(1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right]\right)\left(\frac{\alpha}{\lambda}\left[\frac{N_k H_\beta}{Z_k} + EB^{\beta}_k\right]\right)}{t_k\varepsilon_{NR_k}\varepsilon^{\beta}_{NR_k}}$$

We alternatively manipulate our original two first order conditions to get a price neutral change in the two tax tools. For the tax, we separate out the revenue and avoidance responses

$$\alpha\left[-\frac{\partial q_k}{\partial t_k}x_k + \frac{\partial \Pi_k}{\partial t_k}\right] + \lambda\left[Z_k + t_k\left[\frac{\partial(1-\gamma_k)}{\partial t_k}R_k + (1-\gamma_k)\frac{\partial R_k}{\partial t_k}\right] - \beta_k\frac{\partial R_k}{\partial t_k}\right]$$

We've previously found that

$$\begin{aligned}\frac{\partial R_k}{\partial t_k} &= \frac{\partial x_k}{\partial t_k}q_k + x_k\frac{\partial q_k}{\partial t_k} - \mu\phi_y\frac{\partial x_k}{\partial t_k} \\ &= \frac{\partial q_k}{\partial t_k}x_k\left[1 - \left[\frac{q_k - \mu\phi_y}{q_k}\right]\varepsilon_{x_k}\right]\end{aligned}$$

which means we can convert our expression to

$$\frac{\partial q_k}{\partial t_k}\left[-\frac{\alpha}{\lambda}x_k + x_k\left[1-\left[\frac{q_k-\mu\phi_y}{q_k}\right]\varepsilon_{x_k}\right][(1-\gamma_k)t_k-\beta_k]\right] = -Z_k - t_k\frac{\partial(1-\gamma_k)}{\partial t_k}R_k - \frac{\alpha}{\lambda}\frac{\partial \Pi_k}{\partial t_k}$$

For the enforcement tool, an equivalent process gets us

$$\frac{\partial q_k}{\partial \beta_k}\left[-\frac{\alpha}{\lambda}x_k + x_k\left[1-\left[\frac{q_k-\mu\phi_y}{q_k}\right]\varepsilon_{x_k}\right][(1-\gamma_k)t_k-\beta_k]\right] = R_k - t_k\frac{\partial(1-\gamma_k)}{\partial \beta_k}R_k - \frac{\alpha}{\lambda}\frac{\partial \Pi_k}{\partial \beta_k}$$

Dividing the two equations and rearranging brings us to the expression in the proposition

$$\left[1-\gamma_k+\frac{\frac{\partial q_k}{\partial t_k}}{\frac{\partial q_k}{\partial \beta_k}}\right] - t_k\left[\frac{\partial \gamma_k}{\partial t_k} - \frac{\partial \gamma_k}{\partial \beta_k}\frac{\frac{\partial q_k}{\partial t_k}}{\frac{\partial q_k}{\partial \beta_k}}\right] + \frac{\alpha}{\lambda}\frac{1}{R_k}\left[\frac{\partial \Pi_k}{\partial t_k} - \frac{\partial \Pi_k}{\partial \beta_k}\frac{\frac{\partial q_k}{\partial t_k}}{\frac{\partial q_k}{\partial \beta_k}}\right] = 0$$

Note that we could have instead derived an equation for when an increase in the tax rate and a decrease in the enforcement rate is welfare positive (outside of an optimum). To do so, we differentiate the Lagagrangian with respect to the tax rate without assuming we are at an optimum (not setting the first expression equal to 0). Then we differentiate the Lagrangian with respect to the enforcement rate at an amount that cancels out the price change of the first change (i.e., the ratio of the two pass-throughs). Adding these two expressions together gets us the same above expression, and we can ask when it is greater than or less than 0 (equal to 0 at the optimum).

Since there is no output distortion under a profit tax, we instead add only the RHS of the two FOCs (zeroing out the price related terms) to get

$$[1-\gamma_k+1] - t_k\left[\frac{\partial \gamma_k}{\partial t_k} - \frac{\partial \gamma_k}{\partial \beta_k}\right] + \frac{\alpha}{\lambda}\frac{1}{R_k}\left[\frac{\partial \Pi_k}{\partial t_k} - \frac{\partial \Pi_k}{\partial \beta_k}\right] = 0$$

A.3.4 Corollary 3.1

Under a profit tax ($\mu = 1$), there is no price/output distortion. Then we take our two optimal expressions and simply divide them to get

$$\frac{\frac{t_k}{1-t_k}}{\beta_k} = \frac{\left(1-\frac{\alpha}{\lambda}\right)\left(\frac{R_k}{Z_k} + \frac{\alpha}{\lambda}\frac{1}{Z_k}\frac{H_\beta}{z_k}\right)}{t_k\varepsilon_{\gamma_k}\varepsilon^{\beta}_{\gamma_k}}$$

where the income elasticities were replaced by the avoidance elasticities, and the welfare effects are only due to rent extraction from profits.

A.4 Uniform Taxation

Under uniform taxation, the Lagrangian is

$$\mathcal{L} = v(q) + \lambda \left[\sum_j tZ_j - \beta_j R_j - G \right]$$

Differentiating this expression with respect to the tax rate, we get

$$\sum_j \alpha \left[-\frac{\partial q_j}{\partial t} x_j + \frac{\partial \Pi_j}{\partial t} \right] + \lambda \left[Z_j + t_j \frac{\partial Z_j}{\partial t_j} - \beta_j \frac{\partial R_j}{\partial t} \right] = 0$$

which we can restate as

$$\sum_j \frac{\alpha}{\lambda} \frac{Z_j}{Z_j} \left[-\frac{\partial q_k}{\partial t_k} x_k + \frac{\partial \Pi_k}{\partial t_k} \right] + Z_j - \frac{t_k}{1-t_k} Z_j \varepsilon_{Z_j} - \beta_j \frac{Z_j}{Z_j} \frac{\partial R_k}{\partial t_k} = 0$$

using the same definition of excess burden from before, we can rearrange this expression to get

$$\frac{t}{1-t} = \frac{\sum_j Z_j \left[1 - \frac{\alpha}{\lambda} \left[1 + EB_j \right] - \frac{\beta_j}{Z_j} \frac{\partial R_j}{\partial t} \right]}{\sum_j Z_j \varepsilon_{Z_j}}$$

or, if we include enforcement costs with the taxable income elasticity to get a "net revenue elasticity", we can write this as

$$\frac{t}{1-t} = \frac{\sum_j Z_j \left[1 - \frac{\alpha}{\lambda} \left[1 + EB_j \right] \right]}{\sum_j Z_j \varepsilon_{NR_j}}$$

The enforcement side, however, has the same expression as before since we still allow these to be differentiated. The only change is to replace the differentiated tax rate with the uniform tax rate.

A.4.1 Proposition 4

Suppose we have two industries m and n that equal in all market fundamentals except that n has a higher number of firms. We can divide the two optimal enforcement conditions for a differentiated system to get the ratio of optimal enforcement rates

$$\frac{\beta_m}{\beta_n} = \frac{\frac{H_\beta(r_n,\gamma_n,\beta_n)}{r_n}}{\frac{H_\beta(r_m,\gamma_m,\beta_m)}{r_m}} \frac{\varepsilon^\beta_{\gamma_m}}{\varepsilon^\beta_{\gamma_n}} \frac{(1-\gamma_m)}{(1-\gamma_n)} \frac{t_m}{t_n}$$

Under constant scale avoidance, $\frac{t_m}{t_n} = 1$, so there is no difference in optimal enforcement rates if the tax is forced to be uniform. Suppose instead that at the fully differentiated optimum $\frac{t_m}{t_n} > 1$, which occurs if and only if the avoidance response is higher for industry t_n, i.e, if avoidance costs exhibit decreasing returns to scale. Then if the tax rate is forced to be uniform and this ratio shrinks to 1, it must be the case that the β_n must rise in relation to β_m.

A.5 Heterogenous Firms

Now we assume that the firms in each industry may differ along their productivity margin. For simplicity, assume that we have constant marginal costs and thus firms in an industry are indexed by their marginal cost scalar. Let i index each firm within an industry k. Then the Lagrangian of the government. Expression wise, the optimal formulae will all appear nearly the same since these are written in terms of aggregates. What will change, however, are how these aggregates are composed.

A.6 Competition as a Policy Instrument

In this section, we assume that the government can directly control the number of firms in an industry, i.e., that N_k for each industry is an instrument for the government. Taking the derivative of the Lagrangian with repect to this policy tool, we have that

$$\frac{d\mathcal{L}}{dN_k} = \frac{\alpha}{\lambda}\left[-\frac{\partial q_k}{\partial N_k}x_k + \frac{\partial \Pi_k}{\partial N_k}\right] + t_k\frac{\partial Z_k}{\partial N_k} - \beta_k\frac{\partial R_k}{\partial N_k} - f(N_k)$$

where f is some function that measure the marginal cost of adding another firm. We can think of this as some regulatory action.

First, we examine the welfare effect:

$$\frac{\partial \pi_k}{\partial N_k} = (1-(1-\gamma_k)t_k - H_r)\left[\frac{\partial y_k}{\partial N_k}q_k + y_k\frac{\partial q_k}{\partial N_k}\right] - (1-\mu[(1-\gamma_k)t_k + H_r])\phi_y\frac{\partial y_k}{\partial N_k} - H_\gamma\frac{\partial \gamma_k}{\partial N_k} + t_k\frac{\partial \gamma_k}{\partial N_k}r_k$$

Unlike before, we must make one adjustment to the equilibrium demand and supply condition as the number of firms increases. Particularly, since $x_k = N_k y_k$, we have that

$$\frac{\partial x_k}{\partial N_k} = N_k\frac{\partial y_k}{\partial N_k} + y_k \iff \frac{\partial y_k}{\partial N_k} = \frac{1}{N_k}\left[\frac{\partial x_k}{\partial N_k} - y_k\right]$$

where we did not include the second effect of the changing number of firms due to this being fixed in previous policy tool exercises. We can now rewrite the profit

$$\begin{aligned}
\frac{\partial \pi_k}{\partial N_k} &= (1-(1-\gamma_k)t_k - H_r)\left[y_k\frac{\partial q_k}{\partial N_k} - \frac{\partial q_k}{\partial x_k}\frac{\partial y_k}{\partial N_k}y_k\right] \\
&= (1-(1-\gamma_k)t_k - H_r)\left[y_k\frac{\partial q_k}{\partial N_k} - \frac{\partial q_k}{\partial x_k}\frac{1}{N_k}\left[\frac{\partial x_k}{\partial N_k} - y_k\right]y_k\right] \\
&= (1-(1-\gamma_k)t_k - H_r)\left[y_k\frac{\partial q_k}{\partial N_k}\frac{N_k-1}{N_k} - \frac{\partial q_k}{\partial x_k}y_k^2\right] \\
&= (1-(1-\gamma_k)t_k - H_r)\left[\rho_k^N\frac{N_k-1}{N_k} - \frac{1}{N_k\varepsilon_{x_k}}\right]y_k q_k
\end{aligned}$$

The aggregate profit change is

$$\frac{\partial \Pi_k}{\partial N_k} = (1-(1-\gamma_k)t_k - H_r)\left[\rho_k^N \frac{N_k-1}{N_k} - \frac{1}{N_k \varepsilon_{x_k}}\right] x_k q_k$$

And the price change is

$$\begin{aligned}\left(\frac{\partial \gamma_k}{\partial N_k} t_k - H_{rr}\left[\frac{\partial y_k}{\partial N_k} q_k + y_k \frac{\partial q_k}{\partial N_k} - \mu\phi_y \frac{\partial y_k}{\partial N_k}\right] - H_{r\gamma}\frac{\partial \gamma_k}{\partial N_k}\right)\left[q_k + \frac{\partial q_k}{\partial x_k} y_k\right] \\ (1-(1-\gamma_k)t_k - H_r)\left[\frac{\partial q_k}{\partial N_k} + \frac{\partial^2 q_k}{\partial x_k^2}\frac{\partial x_k}{\partial N_k} y_k + \frac{\partial q_k}{\partial x_k}\frac{\partial y_k}{\partial N_k}\right] \\ -\mu\left(\frac{\partial \gamma_k}{\partial N_k} t_k - H_{rr}\left[\frac{\partial y_k}{\partial N_k} q_k + y_k \frac{\partial q_k}{\partial N_k} - \mu\phi_y \frac{\partial y_k}{\partial N_k}\right] - H_{r\gamma}\frac{\partial \gamma_k}{\partial N_k}\right)\phi_y \\ -(1-\mu[(1-\gamma_k)t_k + H_r])\phi_{yy}\frac{\partial y_k}{\partial N_k} = 0\end{aligned}$$

which we can rearrange as

$$\begin{aligned}\frac{\partial q_k}{\partial N_k}\left[(1-(1-\gamma_k)t_k)\left(1 + \frac{1}{N_k \varepsilon_{ms_k}} + \frac{\varepsilon_{x_k} - \frac{1}{N_k}}{\varepsilon_{s_k}}\right)\right] = \left(-\frac{\partial \gamma_k}{\partial N_k}(1-\mu\phi_y)\left[t_k - H_{r\gamma}\right]\right)\left(q_k + \frac{\partial q_k}{\partial x_k} y_k\right) \\ -\frac{y_k}{N_k}\left[(1-\mu\phi_y)H_{rr}\left[q_k - \mu\phi_y\right]\left[q_k + \frac{\partial q_k}{\partial x_k} y_k\right] - (1-(1-\gamma_k)t_k - H_r)\frac{\partial q_k}{\partial x_k} + (1-\mu[(1-\gamma_k)t_k + H_r])\phi_{yy}\right]\end{aligned}$$

The avoidance response can be obtained by differentiating the avoidance FOC

$$(t_k - H_{\gamma r})\left(\frac{\partial y_k}{\partial N_k} q_k + y_k \frac{\partial q_k}{\partial N_k} - \mu\phi_y \frac{\partial y_k}{\partial N_k}\right) = H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial N_k}$$

which becomes

$$(t_k - H_{\gamma r})\frac{\partial q_k}{\partial N_k}\left(\frac{\partial x_k}{\partial q_k}\frac{q_k}{x_k} + 1 - \mu\frac{\phi_y}{q_k}\frac{q_k}{x_k}\frac{\partial x_k}{\partial q_k}\right) - (t_k - H_{\gamma r})\frac{y_k}{N_k}(q_k - \mu\phi_y) = H_{\gamma\gamma}\frac{\partial \gamma_k}{\partial N_k}$$

A.7 A Pure Profit Tax

Under a truly pure profit tax where all costs are deductible, the firm's profit maximization problem is slightly adjusted as follows

$$\max_{y_k,\gamma_k} y_k q_k(x_k) - C_k(y_k, r_k, \gamma_k, \beta_k) - t_k(1-\gamma_k)(y_k q_k - \mu C_k)$$

where $C = F(y_k) + H(y_k q_k - \mu F - \mu H, \gamma, \beta)$. Then the two adjusted FOC are

$$\begin{aligned} y_k &: (1-(1-\gamma_k)t_k - H_r)\left[q_k + \frac{\partial q_k}{\partial x_k} y_k\right] = (1-\mu[(1-\gamma_k)t_k + H_r])\phi_y \\ \gamma_k &: t_k(y_k q_k - \mu C_k) = H_\gamma - t_k(1-\gamma_k)\mu H_\gamma \end{aligned}$$

where the primary difference can be seen in the avoidance condition. Since avoidance costs are deductible, the marginal cost of avoidance is lowered by the (non-avoided) fraction of avoidance costs. The other difference is that H_r represents a recursive function as $H(\cdot)$ is a function of $H(\cdot)$. Neither of these two adjustments provide substantive differences.