# Optimal Taxation and Enforcement with Market Power

Peter Choi

October 25, 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 several 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 remitting nothing in the previous two years as well.[1] A 2023 Government Accountability Office (GAO) study reported that about a quarter of what they considered profitable large corporations remitted on average zero taxes between 2014-2018, with average effective tax rates hovering around 15% until dropping to single digits due to the 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 hovered around 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 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 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 an inverse 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 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 being versus auditing many smaller firms with an equal aggregate value, regardless of beliefs on underlying avoidance behavior. This administrative cost

[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.

based argument is 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 in particular 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 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 from 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, though 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 primarily bridge these studies with the those that introduce market power into the optimal tax problem (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 the the persistent profitability that market power allows firms the ability 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 relaxes 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 alternations 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 $\Pi_k$ 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 $\gamma_k$, while subject to a tax rate $t_k$ and an enforcement rate $\beta_k$, 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 $\gamma_k$ 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, $\phi$ 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 concentrated 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, results discussing the impact of competition take either a "within market" or an 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.

The firm's problem 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.

Also to note is that this expression implies that

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

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

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

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 Lerner index can be generalized to equal a conduct parameter that nests different models of competition. This paper sets aside this generalization and focuses on the Cournot framework.

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 revenues 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 income of an industry decreases as the number of firm increases, meaning average firm size decreases, we can say that, for a given tax-enforcement combination, the avoidance rate increases as competition increases. 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, we can define an elasticity of the marginal costs of avoidance with respect to firm size

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

An elasticity of 1 represents marginal avoidance costs that scale linearly with size, a special subcase which we will refer 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. We summarize the importance of this relationship 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 noted that term "constant" is on a rate basis. If the rate of avoidance is the same for two differently sized firms, then the larger firm is clearly 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 is neglible industry-specific benefits or access (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 $\beta$ 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 profits), $a(\cdot)$ is the administrative cost function, and $G$ is an exogenous revenue requirement that will ultimately determine the value of the government budget multiplier $\lambda$.[8] Administrative costs are a function of both the effort put in by the government into the industry, $\beta_j$, 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

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

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, $\beta_j$ can directly be interpreted as a rate similar to the tax rate $t_j$. Built into this assumption is the ability to identify $t_j(1-\gamma_j) - \beta_j$ as the mechanical net rate of

[8] A note on the value of $\lambda$ 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, $\lambda$ 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 $\lambda$.

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 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 kept uniform across all industries, though enforcement rates can still be differentiated in line with the assumption that enforcement authorities such as the IRS have relatively more freedom in how they choose to target their resources.

# 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, as behavioral impacts of bundle re-optimization are zeroed out by the envelope condition and income and cross-price effects are assumed zero, and profit responses, since consumers receive all profits from each industry:

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

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

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

On the fiscal side, the base of the tax rates are the reported taxable income 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 rates of each tax tool onto consumer prices are 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.[9]

Likewise, the lack of free entry in imperfectly competitive markets and the (partial) control of pricing implies that pass-through does not necessarily have to be full 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).

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

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

and the enforcement pass-through rate as

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

where $\varepsilon_{x_k}$ is the elasticity of demand and $\varepsilon_{s_k}$ is the elasticity of the inverse marginal cost ("pseudo-supply") curve. The derivation of these pass-through rates follows closely to Weyl and Fabinger

[9] 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 prior includes includes additional changes in net margins due to altering firm size.

(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 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 pseudo-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 scale avoidance, this endogenous avoidance effect drops out. 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. We repeat this in the following lemma.

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

### 3.2 Avoidance Response

We have discussed in Section 2 the implications of the elasticity of marginal avoidance costs on the levels of avoidance. Here we discuss the implications on the behavioral response of avoidance, which is an important component of the reported taxable income response. The change in the avoidance rate in response to an increase in the tax rate is

$$\frac{1}{1-\gamma_k}\frac{\partial \gamma_k}{\partial t_k} = \frac{1}{(1-\gamma_k)H_{\gamma\gamma}}\left[\underbrace{r_k}_{Direct} + \underbrace{t_k(1-\varepsilon^r_{H_\gamma})\frac{\partial r_k}{\partial t_k}}_{\text{Size Response}}\right] \tag{8}$$

which breaks down 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, if $1-\varepsilon^r_{H_\gamma} > 0$, this implies that increasing firm size increases the net margin of avoidance, making further avoidance more attractive.

Combining these terms gets us the sign of the overall 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 firm produces in.

Lastly, under constant scale avoidance, we have that $1 - \varepsilon^r_{H_\gamma} = 0$, indicating no endogenous change in the avoidance margin, 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. This is because

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

under constant scale avoidance. Since $\gamma_k$ 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

$$\frac{1}{1-\gamma_k}\frac{\partial \gamma_k}{\partial \beta_k} = \underbrace{-\frac{H_{\gamma\beta}}{(1-\gamma_k)H_{\gamma\gamma}}}_{Direct} + \underbrace{t_k(1-\varepsilon^r_{H_\gamma})\frac{\partial r_k}{\partial \beta_k}}_{\text{Size Response}} \tag{9}$$

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). There is no direct size related component here, but the baseline level of avoidance matters. 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.

We summarize the relationships between tax and enforcement with market power in the following lemma.

**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

$$\frac{\partial \Pi_k}{\partial t_k} = \underbrace{-Z_k}_{Mechanical} + \underbrace{(1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\rho_k x_k q_k}_{Competition} \tag{10}$$

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

$$\frac{\partial \Pi_k}{\partial \beta_k} = -N_k H_\beta + (1-(1-\gamma_k)t_k - H_r)\frac{N_k-1}{N_k}\rho_k^\beta x_k q_k \tag{11}$$

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 profits. Using our profit response expression, we can write the welfare effect in response to a tax change as

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

Breaking down this expression, the first term is the reported revenue in the industry, which is 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.

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

$$\frac{1}{\alpha Z_k}\frac{dv}{dt_k} = -1 + EB_k \tag{12}$$

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

$$EB_k = \rho_k \frac{x_k q_k}{Z_k}\left[1 - (1 - (1 - \gamma_k)t_k - H_r)\frac{N_k - 1}{N_k}\right] \tag{13}$$

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 our results to obtain the welfare effect per reported revenue on the enforcement side, we have

$$\frac{1}{\alpha Z_k}\frac{dv}{d\beta_k} = -\frac{N_k H_\beta}{Z_k} - EB_k^\beta \tag{14}$$

where $EB_k^\beta$ 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 then excess cost, similar logic can be applied. The first term is the impact on profits per dollar of mechanical tax revenue and is thus our "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

$$EB_k^\theta = \rho_k^\theta\left(1 - \frac{\frac{\phi_y}{q_k}}{1 - \frac{1}{N_k \varepsilon_{x_k}}}\right)\left(1 - \frac{1}{N_k}\right) \tag{15}$$

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, we can say that 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 tax and enforecement levels, as we will see in Section 4.

The most challenging portion of this term to empirically observe are marginal costs 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 can 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 $H_\beta$. There has been recent work in this field 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 tax rate is

$$\frac{\varepsilon_{Z_k}}{1-t_k} = \underbrace{\frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{1}{1-\gamma_k}}_{\text{Avoidance Response}} \underbrace{\frac{(1-\gamma_k)x_k q_k}{Z_k}\rho_k\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]}_{\text{True Taxable Income Response}} \tag{16}$$

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 size response response that can be joined to the second 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

$$\frac{\varepsilon_{Z_k}^{\beta_k}}{\beta_k} = \frac{\partial(1-\gamma_k)}{\partial\beta_k}\frac{1}{1-\gamma_k} + \frac{(1-\gamma_k)x_k q_k}{Z_k}\rho_k^{\beta}\left[1-\left(1-\mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right] \tag{17}$$

To simplify exposition, 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, leaving only the avoidance response. Thus, the relationship between the reported income response and market power exactly follows the relationship between avoidance respose and market power, 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

$$\frac{\varepsilon_{Z_k}}{1-t_k} = \frac{\partial(1-\gamma_k)}{\partial(1-t_k)}\frac{1}{1-\gamma_k} - \rho_k\left[1-\varepsilon_{x_k}\right] \tag{18}$$

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 relates to market power. Consider the constant avoidance technology case. Then 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* $\varepsilon_{x_k} < 1$.

Under non-constant returns avoidance techonology, the "tipping point" of the reported revenue response 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*

$$\frac{t_k}{1-t_k} = \frac{1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right] - \frac{\beta_k}{Z_k}\frac{\partial R_k}{\partial t_k}}{\varepsilon_{Z_k}} \tag{19}$$

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

$$\beta_k = \frac{t_k \varepsilon^{\beta}_{Z_k}}{\frac{R_k}{Z_k} + \frac{\beta_k}{Z_k}\frac{\partial R_k}{\partial \beta_k} + \frac{\alpha}{\lambda}\left[\frac{N_k H_\beta}{Z_k} + EB^{\beta}_k\right]} \tag{20}$$

Again $EB^{\theta}_k$ is representative of the additional distortion taxation (or enforcement in the corresponding enforcement expression) 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 we have discussed 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. We will illustrate in our simulations that this possibility for optimal tax rates (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 indicates that enforcement is highly effective in increasing reported revenue 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 firm 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 indirect 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 we include these administrative costs into 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 relationship becomes 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 optimal tax and enforcement rates are given by*

$$\frac{t_k}{1-t_k} = \frac{1-\frac{\alpha}{\lambda}}{\varepsilon_{\gamma_k}}, \qquad \beta_k = \frac{t_k \varepsilon^{\beta}_{\gamma_k}}{\frac{R_k}{Z_k} + \frac{\alpha}{\lambda}\frac{H_\beta}{z_k}} \tag{21}$$

*Under constant scale avoidance technology, the levels of both the profit tax and enforcement rate are independent of market power. Under increasing (decreasing) 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 is the 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, and only 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, however, is the optimal balance of these two tools, as this accounts for both tools being set optimally. The condition that governs this tradeoff is

$$\underbrace{(1-\gamma_k) - \frac{\partial \gamma_k}{\partial t_k} t_k + \frac{\alpha}{\lambda}\frac{1}{x_k}\frac{\partial \pi_k}{\partial t_k}}_{\text{Marginal Tax Effect}} = \frac{\rho_k}{\rho_k^{\beta}} \underbrace{\left[-1 - \frac{\partial \gamma_k}{\partial \beta_k} t_k + \frac{\alpha}{\lambda}\frac{1}{x_k}\frac{\partial \pi_k}{\partial \beta_k}\right]}_{\text{Marginal Enforcement Effect}} \tag{22}$$

or, alternatively stated,

$$\underbrace{\left[(1-\gamma_k) + \frac{\rho_k}{\rho_k^{\beta}}\right]}_{\text{Mechanical Change}} + \underbrace{\left[-\frac{\partial \gamma_k}{\partial t_k} + \frac{\partial \gamma_k}{\partial \beta_k}\frac{\rho_k}{\rho_k^{\beta}}\right] t_k}_{\text{Behavioral Change}} + \underbrace{\frac{\alpha}{\lambda}\frac{1}{x_k}\left[\frac{\partial \pi_k}{\partial t_k} - \frac{\partial \pi_k}{\partial \beta_k}\frac{\rho_k}{\rho_k^{\beta}}\right]}_{\text{Profit Change}} = 0 \tag{23}$$

While the above expression speaks to the optimum, we can also use the LHS of the same expression 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 $\frac{\alpha}{\lambda} < 1$. We can equivalently demonstrate this tradeoff by combining our two expressions from Proposition 2.

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

$$\frac{\frac{t_k}{1-t_k}}{\beta_k} = \frac{\left(1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right]\right)\left(\frac{\alpha}{\lambda}\left[\frac{N_k H_\beta}{Z_k} + EB^{\beta}_{k}\right]\right)}{t_k \varepsilon_{NR_k} \varepsilon^{\beta}_{NR_k}} \tag{24}$$

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

This result is exactly 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 our 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, we can directly state 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 get 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 avoidance technology, this ratio decreases (increases) as the market power increases.*

As we can see by combining our profit tax and pass-through expressions from Section 3, the profit change under constant 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 our results from Section 3, the avoidance responses and the pass-through ratio are both independent of firm size. Therefore, the three terms are all independent of firm size and therefore the optimal relationship between taxation and enforcement is independent of market power. We combine these statements 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 $\frac{t_k^R}{\beta_k^R}$ represent the ratio of revenue maximizing rates. If the market has firms with constant scale avoidance technology, then*

$$\frac{t_k^*}{\beta_k^*} = \frac{t_k^R}{\beta_k^R}$$

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

## 5 Simulations

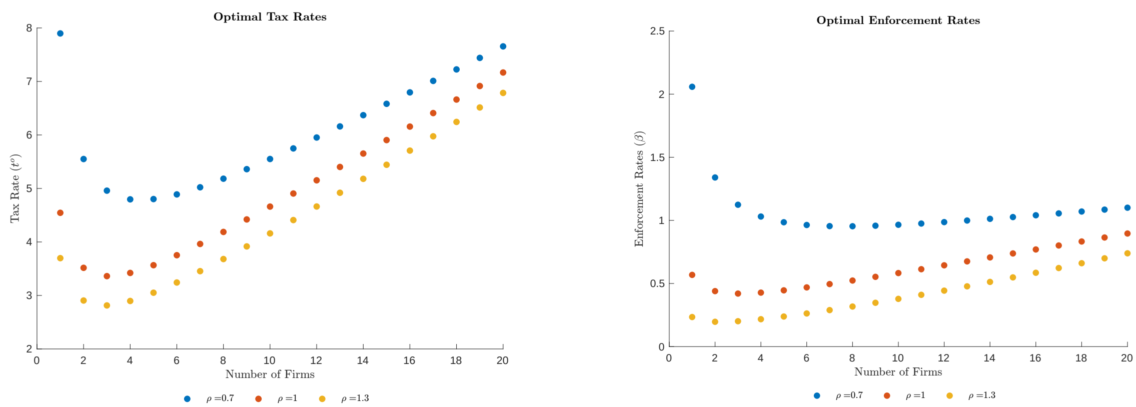
To provide better intution into how the structure of the cost function can affect the results, it is fruitful to illustrate some of these findings with a more concrete cost function. Consider the specifications of the form

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

where $H(r)$ will be of the form $r^\rho$. The parameter $\rho$ is then a measure of increasing (cost-reducing) complementarity between the avoidance rate and output. In this specification, $\rho = 1$ is equivalent to a elasticity of marginal avoidance costs $\varepsilon^r_{H_\gamma} = 1$, and a higher (lower) $\rho$ increases (decreases) this elasticity. $\beta$, 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 $\gamma$ is squared to allow for strictly convex evasion costs.

## 5.1 Simulation Results: Output Tax

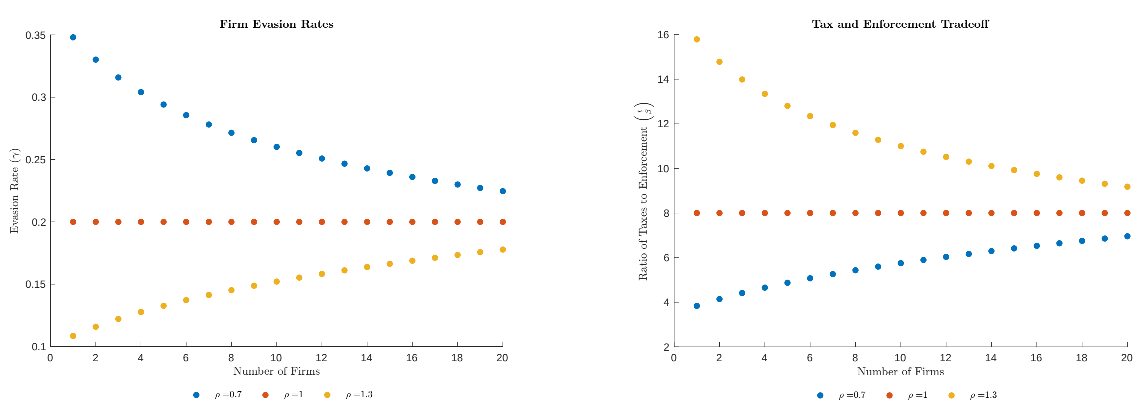
Here we describe the results of simulations of the model presented above. For the starting simulation environment, we examine variants of a single industry ("within market" comparison) and employ a linear demand curve of the form $q(x) = A - Bx$, and the format of the cost function is described as above. The cost function is further parameterized with squared direct production costs, $F(y) = Ky^2$, and a power form for size in the avoidance cost, $G(y) = (yq)^\rho$. 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, representative 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




(a) (b)

Figure 2: Avoidance Rates and Policy Tradeoff

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 $\varepsilon_{x_k}$ transitions from a value above one to a value below 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 production 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 production 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.

We can see this figure inversely mirrors Figure 2a on avoidance rates. This aligns with our discussions on the relationship between the tax and enforcement bases as a function of the avoidance rates. When $\rho$ is high, the elasticity of marginal costs of evasion 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 other effects are not large enough to overturn this primary effect.

### 5.1.1 Elasticity of Demand [In Progress]

The simulations thus far have been under the assumption of a linear demand function. However, there are several ways the assumption on the convexity of the demand function may affect the problem. For example, it is well known in the literature that a linear demand function implies an inverse relationship between tax pass-through and competition since linear demand falls under the log-concave classification. Under other log-convex demand functions, however, it is also possible that the reverse is true.

## 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

$$\frac{t}{1-t} = \frac{\sum_j Z_j \left[1 - \frac{\alpha}{\lambda}\left[1 + EB_j\right] - \frac{\beta_j}{Z_j}\frac{\partial R_j}{\partial t}\right]}{\sum_j Z_j \varepsilon_{Z_j}} \tag{25}$$

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

$$\frac{t_k}{1-t_k} = \frac{1-\frac{\alpha}{\lambda}}{\varepsilon_Z} \tag{26}$$

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, we can divide our optimal enforcement condition from before with the same condition for a different market, i.e.,

$$\frac{\beta_m}{\beta_n} = \frac{\frac{H_\beta(z_n,\gamma_n,\beta_n)}{z_n} + EB_n}{\frac{H_\beta(z_m,\gamma_m,\beta_m)}{z_m} + EB_m} \frac{\varepsilon^\beta_{NR_m}}{\varepsilon^\beta_{NR_n}} \tag{27}$$

which follows a similar comparison across industries as before. The key difference is that at the optimum, this ratio does not account for the ratio in differentiated tax rates. Under the profit tax, this is even simpler

$$\frac{\beta_m}{\beta_n} = \frac{\frac{H_\beta(z_n,\gamma_n,\beta_n)}{z_n}}{\frac{H_\beta(z_m,\gamma_m,\beta_m)}{z_m}} \frac{\varepsilon^\beta_{Z_m}}{\varepsilon^\beta_{Z_n}} \tag{28}$$

With differentiated taxes, this term 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 $\beta_m$ relative to $\beta_n$ to compensate. If the ratio is below 1, then we have the inverse case. 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 than it would have been in a world with a differentiated profit tax.*

We can explain this imperfect instruments logic as follows. 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 all 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.

# 7 Extensions and Alternative Specifications of the Model

In this section, we consider a few extensions and alternative specifications of the model.

## 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 revenues. Thus, this version of the model implicitly assumes that they are able to observe both true revenues 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

$$\sum_j t_j Z_j - \eta(N_j)a(\beta_j) \tag{29}$$

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 our optimum expressions,

$$\frac{t_k}{1-t_k} = \frac{1 - \frac{\alpha}{\lambda}\left[1 + EB_k\right]}{\varepsilon_{Z_k}} \tag{30}$$

and

$$\beta_k = \frac{t_k \varepsilon^{\beta}_{Z_k}}{\frac{\alpha}{\lambda}\left[1 + EB^{\beta}_k\right] + \frac{\eta(N_k)}{Z_k} a_{\beta}} \tag{31}$$

In the tax expression, there is no longer an effect on the tax rate due to the true revenue changing sizes affecting total enforcement costs. In the enforcement expression, the administrative cost term appears in the denominator, divided by reported revenue. 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

$$r_k H(N_k, \gamma_k, \beta_k) \tag{32}$$

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

$$t_k = H_\gamma(N_k, \gamma_k, \beta_k) \tag{33}$$

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.

### 7.3 Heterogenous Firms [In Progress]

Suppose that the firms within a given industry are heterogeneous in their productive capability, i.e., different levels of marginal costs. Ultimately, the expressions should remain the same in formulation, though the underlying substance may change. Both elasticities (with respect to either tax and enforcement) of reported taxable income and the excess burden terms are weighted averages of individual firm responses where weights are not equal as in the homogeneous case.

The details of these changes can be found in the appendix.

### 7.4 Differential Welfare Weights on Profits [In Progress]

We have thus far ignored any consideration of redistribution in our problem. One manner in which we may simply incorporate some degree of redistribution is to put a different weight on welfare associated with profits and those associated with consumer surplus. Let $\zeta$ represent the weight on profits, such that the welfare effect of a tax tool change is

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

where $\zeta < 1$ would indicate profits (producer surplus) are valued less than consumer surplus. Since there is no consumer surplus effect under the profit tax case, there is ultimately no change to the previous analysis as all forces work solely through profits. For non-full deductibility, however, this is no longer the case.

### 7.5 Market Power as a Choice Variable [In Progress]

An alternative potential instrument to the government in this problem is directly addressing market power via regulatory action. A potentially simple way to illustrate this idea in the context of the current model is to have $N_k$ be a control variable for the government. Increasing $N_k$ is to increase the competition within an industry (at some regulatory cost).

## 8 Discussion and Conclusion

Whether market concentration and 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. This is because higher avoidance rates imply a relatively low tax base as compared to the enforcement 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 latter 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 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 A Pure Profit Tax

Under a truly pure profit tax where all costs are deductible, neither the production or avoidance margins are distorted. To see this, cosider the maximization problem for the firm

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

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

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

## A.2 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 us with these responses, we recall our first order conditions for production and for avoidance:

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

We can also rearrange the first production condition in terms of 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.2.1 Profit Responses

We differentiate the firm's after-tax profits $\pi_k$ with respect to the tax rate:

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

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 ($\Pi_k$) 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

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

### A.2.2 Proof of Lemma 1

Under Assumption 1, we have that the optimal choice in the avoidance rate satisfies

$$I(\gamma_k) = \frac{r_k}{G(r_k)}\frac{t_k}{(1-\mu t_k)J(\beta)}$$

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

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

where $I_\gamma^{-1}$ is a monotonic increasing function (since we assumed $I_\gamma$ is). Clearly, this implies that if $G(\cdot)$ is strictly convex, and thus the denominator inside the parentheses increases faster than the numerator, then $\gamma$ decreases as individual firm size increases. This occurs, in our Cournot framework, when the number of firms decreases, 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.

#### A.2.3 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:

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

which we can rearrange as

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

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

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

The first line is zeroed since the second bracketed term is zero by the production condition. Then, we are left with

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

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. We can find the same consequence for the pass-through rate of the enforcement tool.

Returning to the case where $\mu < 1$, to simplify this expression, we first define a few elasticities. 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 we have that 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, we have that 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

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

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

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

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

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

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

### A.2.4 Proof of Lemma X

The derivation follows closely to Adachi and Fabinger (2022). Suppose that the firm faces both an excise ($v_k$) and an ad valorem tax $t_k$ on output. Then the production FOC is

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

The effective marginal cost to the firm is

$$\tilde{MC}_k=\frac{\phi_y+v_k(1-\gamma_k)}{1-(1-\gamma_k)t_k-H_r}$$

Consider changes in each tool to keep this effective marginal cost constant

$$\frac{\partial}{\partial t_k}\left(\frac{\phi_y+v_k(1-\gamma_k)}{1-(1-\gamma_k)t_k-H_r}\right)dt_k+\frac{\partial}{\partial v_k}\left(\frac{\phi_y+v_k(1-\gamma_k)}{1-(1-\gamma_k)t_k-H_r}\right)dv_k=0$$

which becomes

$$\left(\frac{(1-(1-\gamma_k)t_k-H_r)(-v_k\frac{\partial\gamma_k}{\partial t_k})+(1-\gamma_k-\frac{\partial\gamma_k}{\partial t_k}(t_k-H_{r\gamma}))(\phi_y+v_k(1-\gamma_k))}{(1-(1-\gamma_k)t_k-H_r)^2}\right)dt_k$$
$$+\left(\frac{(1-(1-\gamma_k)t_k-H_r)(1-\gamma_k-\frac{\partial\gamma_k}{\partial v_k}v_k)+(\phi_y+v_k(1-\gamma_k))(\frac{\partial\gamma_k}{\partial v_k}(t_k-H_{r\gamma}))}{(1-(1-\gamma_k)t_k-H_r)^2}\right)dv_k=0$$

which simplifies to

$$\left(\frac{(1-(1-\gamma_k)t_k-H_r)(-v_k\frac{\partial\gamma_k}{\partial t_k})+(1-\gamma_k-\frac{\partial\gamma_k}{\partial t_k}(t_k-H_{r\gamma}))(\phi_y+v_k(1-\gamma_k))}{(1-(1-\gamma_k)t_k-H_r)^2}\right)dt_k$$
$$+\left(\frac{(1-(1-\gamma_k)t_k-H_r)(1-\gamma_k-\frac{\partial\gamma_k}{\partial v_k}v_k)+(\phi_y+v_k(1-\gamma_k))(\frac{\partial\gamma_k}{\partial v_k}(t_k-H_{r\gamma}))}{(1-(1-\gamma_k)t_k-H_r)^2}\right)dv_k=0$$

Clearly this does not work. Instead, if we did things in terms of "effective tax rates" from the firm's perspective, i.e., $\tau_k=(1-\gamma_k)t_k+H_r$ and $\nu_k=(1-\gamma_k)t_k$. We get that

$$\frac{\partial}{\partial\tau_k}\left(\frac{\phi_k+\nu_k}{1-\tau_k}\right)d\tau_k+\frac{\partial}{\partial\nu_k}\left(\frac{\phi_k+\nu_k}{1-\tau_k}\right)d\nu_k=0$$

which will then lead to the exact same conclusion and Adachi and Fabinger (2022).

### A.2.5 Avoidance Response

For the avoidance response, we differentiate the avoidance first order condition with respect to our two tax tools. As before, we start with the tax rate.

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

which can be rearranged as

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

or

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

On the enforcement side, we have that

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

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, taking the avoidance FOC, we have that

$$t_kr_k=H_\gamma$$

Using our definition of the elasticity of marginal evasion costs, we have 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.2.6 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, we plug in our value for the profit change calculated previously to get

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

We can define the "1" as the mechanical dollar raised, while the rest of the expression is the excess burden. Thus, we can simplify this expression as

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

Similarly, for the enforcement, we have that 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 we can equivalently define an excess burden type term

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

As we have previously shown, under a pure profit tax where $\mu = 1$, then the price pass-through is zero for both tool, and therefore, we have that $EB_k = EB^\beta_k = 0$. Thus, the per taxable income welfare effect of taxation is independent of competition (and firm size). Meanwhile, the welfare effect of enforcement is dependent on

### A.2.7 Taxable Income Effect

The relevant elasticity is the elasticity of taxable income, which is reported revenue is the 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, and thus, we have that

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

i.e., that the elasticity of reported taxable income is exactly equal to the tax elasticity of avoidance rate. Similarly, for our elasticity of enforcement, we have

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

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

$$\varepsilon^{\beta_k}_{Z_K} = \varepsilon^{\beta_k}_{1-\gamma_k}$$

Returning to our tax expression, if we plug in our expression for the evasion response, we

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

## A.3 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.3.1 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

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

### A.3.2 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

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

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

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

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

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

### A.3.3 Reported Income

## A.4 Core Results

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

### A.4.1 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

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

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

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

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

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

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

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

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.

### A.4.2 Proof of Corollary X1

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

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

Note that while we rewrote the behavioral elasticity to just be the avoidance elasticity to highlight the lack of production based distortion, the sufficient statistic would still be the elasticity of taxable income.

### A.4.3 Proof of Corollary 2.1

Under constant scale avoidance, we have that $\varepsilon_{\gamma_k}$ is independent of the market power within a market. Thus, the tax rate does not depend on market concentration. Similarly, $\varepsilon_{\gamma_k}^{\beta}$ is independent of market power. 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.

### A.4.4 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.

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

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

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

For the enforcement tool, an equivalent process gets us

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

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.4.5 Proof of Corollary X2

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

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

Again, outside of the optimum, we can simply ask when the above is greater than 0 for a tax-increase, enforcement decrease policy change.

### A.5 Firm Costs, Avoidance, and Relation to Market Power

Of central importance to this paper is the relationship between firm size (or probability in the case of the profit tax) and avoidance in the firm's cost function. To examine this in more detail, again consider the firm's avoidance FOC

$$t_kr_k=(1-\mu t_k)H_\gamma(r_k,\gamma_k,\beta_k)$$

For expositional simplicity, we will consider the class of multiplicatively separable cost functions

**Assumption 1.** *Avoidance costs to the firm are multiplicatively separable such that $H(r_k,\gamma_k,\beta_k)=G(r_k)I(\gamma_k)J(\beta_k)$. $I(\cdot)$ is a strictly increasing monotonic, twice continuously differentiable, strictly convex function.*

#### A.5.1 Proof of Lemma X2

Total avoidance costs for an optimally chosen leve of avoidance are given by

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

And total costs per dollar of firm size ("average cost" of sorts)

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

Let $\nu\equiv\frac{r_k}{G(r_k)}$. Then the relationship between average costs and firm size will be determined by the functional form of

$$\frac{1}{\nu}I(I_\nu^{-1}(D\nu))$$

where $D$ is just some constant. We know that I is montonic increasing and strictly convex, and $I_\nu^{-1}$ is monotonic strictly increasing. Then $I(I_\nu^{-1}(D\nu))$ is convex in $\nu$. Thus, we can conclude that the above expression is weakly proportional to $\nu$ (it may be constant). Thus, when firm size facilitates avoidance, average costs weakly increase. Note the same relation holds for average marginal costs of enforcement activity on the firm, i.e., $\frac{H_\beta}{r_k}$.

## A.6 Comparative Statics on Market Power

We first reexamine our elasticity of reported revenue, but paricularly focus on the true income response:

$$\frac{\varepsilon_{R_k}}{1-t_k} = \frac{1}{q_k}\frac{\partial q_k}{\partial t_k}\left[1 - \left(1 - \mu\frac{\phi_y}{q_k}\right)\varepsilon_{x_k}\right]$$

Under constant scale avoidance, this is the only term that matters for determining how the reported income elasticity changes with respect to market power. To derive intuition, we first consider how this changes under an output tax $\mu = 0$. Differentiating this condition, we get

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

Under log-concave demand, the pass-through rate increases with the number of firms, meaning that the first term is positive when $\varepsilon_{x_k} < 1$ and negative when $\varepsilon_{x_k} > 1$.

### A.6.1 Proof of Proposition X3

*Proof.* We differentiate the elastity of reported revenue with respect to the number of firms. Under constaint 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 is

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

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. For log-convex demand functions, the pass-through rate is decreasing with number of firms, and the demand elasticity is increasing with number of firms. As long as $\varepsilon_{x_k} < 1$, both terms are positive and thus the elasticity of reported revenue is larger. □

### A.6.2 Proof of Proposition X4

*Proof.* We differentiate the elastity of reported revenue with respect to the number of firms. Under constaint 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 is

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

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. The sign of the last term depends on the firm revenue in the denominator. Since the revenue change with respect to the number of firms is given by

$$\frac{d}{dN}x_kq_k=\frac{\partial q_k}{\partial N_k}(1-\varepsilon_{x_k})x_k$$

The price should decrease with the number of firms, and therefore if $\varepsilon_{x_k}<1$, the total revenue should decrease. Since this term is in the denominator, the overall change of this term is positive. Thus, all three terms are positive and the overall expression is positive, meaning the elasticity of reported revenue is larger. For log-convex demand functions, the pass-through rate is decreasing with number of firms, the demand elasticity is increasing with number of firms. However, the last term is still positive and we cannot explicitly sign the overall expression. □

## A.7 Uniform Taxation

Under uniform taxation, the Lagrangian is

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

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.7.1 Corollary Y

Suppose we have two industries $m$ and $n$. Then we can divide the two optimal enforcement conditions to get the relative ratio of enforcement

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

When taxes are differentiated, this expression is multiplied by $\frac{t_m}{t_n}$

## A.8 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. We will have to make a few adjustments.

## A.9 Alternative Enforcement Costs

In this section, we consider several alternatives to the enforcement costs specified in the baseline model which are directly related to the (true) size of the industry. Again, this specification implicitly assumes that the government knows in advance or has perfect expectations on the degree of avoidance in all industries, which is how they are able to balance their budget. We can specify alternative versions of the model that do not require this assumption.

### A.9.1 Per-Firm Enforcement Cost

## A.10 Assumptions on the Cost Function

Note that the Hessian of the profit function of a monopoly is as follows

$$H = \begin{bmatrix} \frac{\partial q_k}{\partial y_k} + \frac{\partial^2 q_k}{\partial y_k^2} y_k + \frac{\partial q_k}{\partial y_k} - C_{yy} & t_k - C_{y\gamma} \\ t_k - C_{y\gamma} & -C_{\gamma\gamma} \end{bmatrix}$$

Then the determinant is

$$-\left[\frac{\partial^2 q_k}{\partial t_k^2} y_k + 2\frac{\partial q_k}{\partial t_k} - C_{yy}\right] C_{\gamma\gamma} - (t - C_{x\gamma})^2$$

In order to be at an optimum, it must be true that this determinant is greater than 0, and thus it be be the case that

$$(t - C_{x\gamma})^2 < -\left[\frac{\partial^2 q_k}{\partial t_k^2} y_k + 2\frac{\partial q_k}{\partial t_k} - C_{yy}\right] C_{\gamma\gamma}$$

where the right side is positive since $C_{\gamma\gamma} > 0$ and the term in the brackets is $< -0$ in order to be a maximum.

## A.11 Monopoly Taxation: A Pigouvian Argument?

In this section, we discuss the relationship between "correcting" the monopoly problem and correcting an externality via a typical Pigouvian tax. As a monopoly leads to underproduction of a good, the naive equivalent would be a positive extenernality such that there is a gap between the private marginal benefit and social marginal benefit. To make this comparison we consider the following exercise: For a given monopoly market we find an equivalent external marginal benefit in a competitive market that would make it such that the level of production is the same across the two markets. In other words, the demand function in the monopoly market is equal to the social marginal benefit curve in the externality market. Next, we find the subsidy in each market that would close this wedge and return each market to the social equilibrium. In the externality market, this is simply equal to the marginal benefit at the point of production. The exercise then is to see whether the same is true for the monopoly market.

Consider a monopoly market producing a quantity $x_m$, where $x_m$ must satisfy the profit condition equating marginal benefit and cost

$$q_m + \frac{\partial q_m}{\partial x_m} x_m = C_y$$

Let $\xi_b$ represents the marginal externality benefit in the competitive market. The social marginal benefit is the $q_c + \xi_b$, and therefore the private equilibrium is governed by the condition

$$q_c + \xi_b = C_y$$

Thus, it should be clear that the equivalent marginal externality benefit that makes the competitive case ex ante similar to a monopoly is

$$\xi_b = \frac{\partial q_m}{\partial x_m} x_m$$

Now suppose that the government can impose a tax in each market to restore output to the socially competitive level. As stated before, if $t_c = -\xi_b$, then the competitive condition becomes

$$p_c = C_y$$

which would mean that the competitive industry is producing exactly where price equals marginal costs (inverse supply) and so producing efficiently. Is the same true for the monopoly industry? Not quite. This is because the monopoly will endogenously react to the tax rates, and so the equivalent marginal damage changes along the pathway back to the social equilibrium point. When applying the same tax rate as in the externality problem, the monopoly's first order condition becomes

$$p_m(x'_m) + \frac{\partial q_m}{\partial x'_m} x'_m - \frac{\partial q_m}{\partial x_m} x_m = C_y(x'_m) \iff p_m(x'_m) - C_y(x'_m) = \frac{\partial q_m}{\partial x_m} x_m - \frac{\partial q_m}{\partial x'_m} x'_m$$

Notice that $x'_m$ is *not* the output which equates price and marginal cost, meaning it is not the socially optimal output. It misses the mark by the difference between the marginal surplus between the two points. So what is actual subsidy needed? It must be the case that the marginal revenue at this output equals the net marginal cost. The output, of course, should be $x'_c$ where $x'_c$ is the externality corrected output in the competitive side. Thus, we need that

$$p_m(x'_c) + \frac{\partial q_m}{\partial x'_c} x'_c = C_y(x'_c) + s_m \iff s_m = p_m(x'_c) + \frac{\partial q_m}{\partial x'_c} x'_c - C_y(x'_c)$$

Well the price should equal marginal cost at the social optimum and thus, it must be the case that

$$s_m = \frac{\partial q_m}{\partial x'_c} x'_c$$

Thus, the required subsidy to push the monopoly to the socially competitive level of output is the marginal surplus at this social optimum. Thus the difference between the monopoly correction and the equivalent Pigouvian correction is

$$s_m - \xi_b = \frac{\partial q}{\partial x}(x'_c) x'_c - \frac{\partial q}{\partial x}(x_m) x_m$$