Three Essays on Optimal Policy

by

Daniel Schaffa

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2017

Doctoral Committee:

Professor James R. Hines Jr., Chair
Professor Paul N. Courant
Professor James Jondall Prescott
Professor Joel B. Slemrod

dschaffa@umich.edu

ORCID: 0000-0002-4677-1463

ACKNOWLEDGEMENTS

I am deeply grateful for the continuous guidance and counsel provided to me by my dear friend, Jim Hines. I thank my dissertation committee, Paul Courant, Jim Hines, JJ Prescott, and Joel Slemrod, for useful feedback and helpful advice. Several other economists have imparted upon me insights that have changed my worldview and have opened their doors to me when I required assistance, including John Bound, Martha Bailey, John DiNardo, Chris House, Mike Mueller-Smith, Elyce Rotella, Jeff Smith, Gary Solon, and Frank Thompson—I thank them all. I am also grateful to many professors at the University of Michigan Law School for useful conversations and advice—these include Michael Barr, John Hudson, Kyle Logue, Bill Miller, Bill Novak, John Pottow, Don Regan, and Veronica Santarosa.

Much of this dissertation is based on work coauthored with Dan Jaqua. His interest, skills, and efforts have been a tremendous benefit to this research. Collaborating with him has been rewarding to me professionally and personally.

I have learned so much from my fellow graduate students. I am especially indebted to Jacob Bastian, Ariel Binder, Will Boning, Eric Chyn, Austin Davis, Adam Dearing, Max Farrell, Ed Fox, Alan Griffith, Morris Hamilton, Enda Hargaden, Yeliz Kacamak, Jason Kerwin, Andrew Litten, Ben Meiselman, Christian Proebsting, Daniel Reck, Dimitrije Ruzic, Nate Seegert, Isaac Sorkin, Bryan Stuart, Brenden Timpe, Eleanor Wilking, and Mike Zabek.

The staff at the University of Michigan Department of Economics and the Population Studies Center have done a fantastic job making the ardors of the doctoral program seem but a trifle. I would especially like to express my gratitude to Mary Braun, Laura Flak, Heather MacFarland, Olga Mustata, and Lauren Pulay.

Various iterations of the works included in this dissertation were presented at the University of Michigan Economics Department, the University of Michigan Law School, the University of South Carolina School of Law, the University of Miami School of Law, the University of Richmond School of Law, the George Mason University School of Law, the 2015 Oxford University CBT Doctoral Meeting, the NTA 2016 Annual Conference on Taxation, the 2017 Mid-Atlantic Junior Faculty Forum, and the 2017 ALEA Annual Meeting. I thank the audiences for insightful questions and comments.

I gratefully acknowledge support from the NIA training grant to the Population Studies

Center at the University of Michigan (T32 AG000221), from Rackham Graduate School, and from the Economics Department at the University of Michigan.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

The Case for Subsidizing Harm: Second-best Pigouvian Taxation with Multiple Externalities

From a work with Daniel Jaqua

Abstract

Most energy production activities are subsidized despite generating negative externalities. We explain this phenomenon by developing a model that generalizes previous work on second-best Pigouvian taxation. In this model the policymaker will optimally subsidize a harmful production activity if a constraint or cost prevents the first-best correction of an even more harmful alternative. We highlight three examples. First, it may be optimal to subsidize a harmful activity if a political constraint prevents the taxation of an even more harmful substitute. Second, it may be optimal to subsidize a harmful activity if there is a large administrative cost associated with taxing an even more harmful substitute. Third, it may be optimal to subsidize a harmful production process if the activity mix at lower levels of output uses more harmful activities than the activity mix at higher levels of output.

Keywords: Administrative cost, Externality, Optimal tax, Optimal tax systems, Pigouvian tax, Second-best
JEL Codes: H21, H23

Taxes used to correct externality-generating behaviors are named for Arthur Cecil Pigou, who first described many of their features (Pigou, 1920). Pigouvian taxes improve welfare by aligning private incentives to a notion of public wellbeing. A broad class of policies that influence behavior—including carbon taxes, gasoline taxes, and toll roads—fit into a Pigouvian tax framework.

Classical Pigouvian analysis directs policymakers to tax production activities that generate external harm, setting the tax equal to the marginal harm. In practice many harmful activities are untaxed and surprisingly some are subsidized. This is particularly apparent in energy production as the following table illustrates.[1]

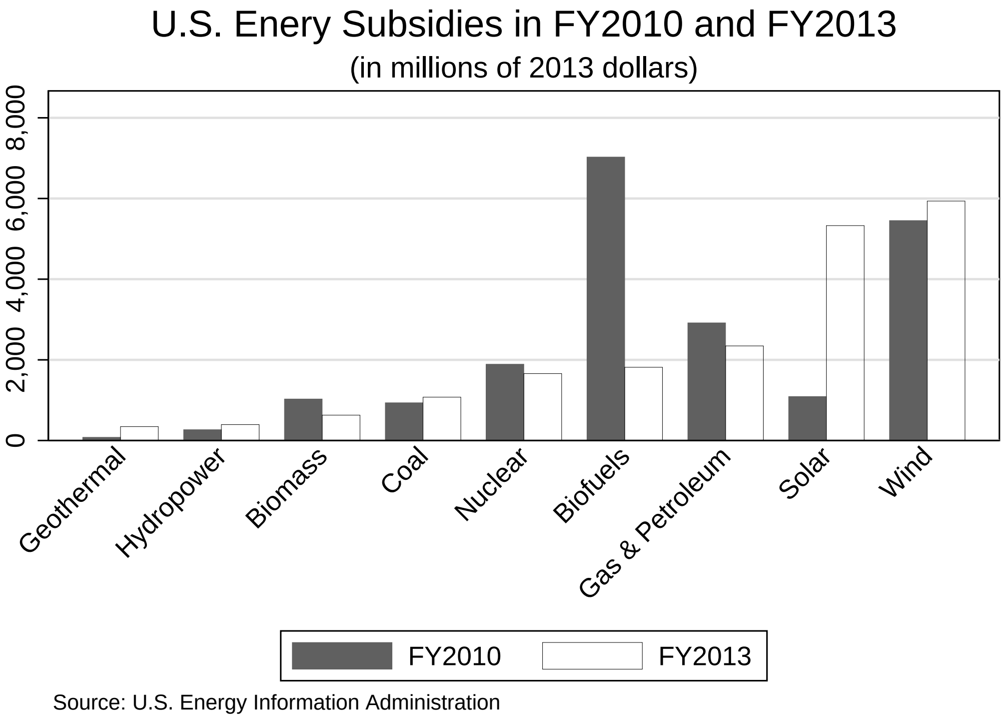
Table 1.1 Energy technologies, externalities, and policies

Technology	Externalities	U.S. Policy
Coal	Greenhouse gases, Acid rain, Hazardous waste, Airborne particulates, Risk of mining accidents	Tax & Subsidize
Oil	Greenhouse gases, Hazardous waste, Airborne particulates, Risk of drilling accidents, Risk of oil spills	Tax & Subsidize
Natural Gas	Greenhouse gases, Ecosystem destruction, Airborne particulates, Water contamination	Tax & Subsidize
Nuclear	Hazardous waste, Risk of nuclear meltdown	Subsidize
Hydropower	Destruction of fisheries and environments, Risk of dam failure	Subsidize
Bioenergy	Greenhouse gases	Subsidize
Solar	Toxic production process, Ecosystem destruction	Subsidize
Geothermal	Toxic gases	Subsidize
Wind	Toxic production (batteries), Eyesore, Harm to wildlife	Subsidize

While this table abstracts from many of the nuances that underly these policies, it clearly demonstrates that classical (or first-best) Pigouvian analysis is insufficient to explain ob-

[1] This is a partial list of the external harms discussed in the World Energy Council's 2016 World Energy Resource Report.

served policy. Moreover the subsidies related to energy production exceed $400 billion per year worldwide and substantially lower the cost of many different energy production technologies as the following graph illustrates.



This paper resolves the apparent discrepancy between theory and practice by placing the Pigouvian taxation of multiple externalities in the second-best (Lipsey and Lancaster, 1956). In doing so, this paper becomes the first to suggest several cases in which an optimal tax system would subsidize externally harmful activities.[2]

Relevant literature. Most of the research on Pigouvian taxes in the second-best has explored how a policymaker should use corrective taxation in the presence of other distortionary taxes (Bovenberg and van der Ploeg, 1994; Bovenberg and Goulder, 1996; Pirttilä and Tuomala, 1997; Pirttilä, 2000; Cremer and Gahvari, 2001; Gahvari, 2014; Jacobs and de Mooij, 2015). This paper takes as a premise that a corrective tax may not be implementable in its first-best form even in the absence of distortionary taxes.

Some work has explored inherently second-best Pigouvian taxation.[3] For example, Polinsky and Shavell (1982) consider Pigouvian taxes with one externality and administrative cost. Fullerton and West (2002) study the indirect taxation of a single externality, specifically when emissions are not taxable, but the policy maker can tax gasoline consumption and

[2] Bovenberg and Goulder (1996) suggest that a policymaker might subsidize a harmful activity if other taxes are set suboptimally.

[3] A parallel literature studies criminal sanctions that are costly to administer and enforce (Polinsky and Shavell, 1992; Kaplow, 1990*a*,*b*).

certain attributes of vehicles. Fullerton and Wolverton (2005) examine optimal policy when a two-part instrument must be used in lieu of a direct tax on a harmful activity. Jacobsen et al. (2016) use a sufficient statistics approach that enables comparison between imperfect corrective tax policies.

This paper generalizes existing work, allowing both direct and indirect taxation and accounting for a wide range of constraints and costs that would push analysis to the second-best. More importantly, this paper extends existing work by allowing multiple externalities in the second-best. In this world, the policymaker must adhere to policy constraints while minimizing lost private benefit, external harm, and the administrative cost of taxation. The policymaker must choose the optimal tax system (Slemrod, 1990; Slemrod and Yitzhaki, 2002), selecting both the optimal tax base and the optimal tax rates (Yitzhaki, 1979). Following Mayshar (1991) each implemented tax instrument is used up to the point where its marginal benefit begins to be eclipsed by its marginal harm. A central feature of our model is that when first-best policy is unavailable, the complementarity between different activities matters (Corlett and Hague, 1953; Sandmo, 1978; Cremer and Gahvari, 1993; Cremer, Gahvari and Ladoux, 1998; Cremer and Gahvari, 2001). In plausible settings the policymaker will optimally subsidize a harmful activity because a constraint or cost prevents the first-best correction of an even more harmful alternative.

Roadmap. The next section introduces a simple and flexible model with multiple production activities each of which may have an associated externality. Unsurprisingly, the first-best Pigouvian tax on each activity is equal to the marginal external harm from each activity. The remaining sections explore Pigouvian taxation in the second-best, emphasizing novel scenarios in which the policymaker would optimally subsidize a harmful activity.

Section two describes how the optimal tax changes when the social planner faces constraints. The social planner is first constrained to tax only a subset of production activities. Incomplete taxation would result if the policymaker were unaware of all production activities or if a powerful lobby prevented taxes on certain activities. If a very harmful activity cannot be taxed, then it may be optimal to subsidize a less harmful substitute.

The social planner is then constrained to tax only output. Output taxation would result if the social planner could collect data on market transactions but not production activities. Even if all production activities were harmful, the social planner should subsidize output if the activity mix at lower levels of output uses more harmful activities than the activity mix at higher levels of output.[4]

Section three describes how the optimal tax changes when taxes are administratively

[4] This generalizes Plott (1966) which shows that optimal direct and indirect taxes may have opposite signs if a single harmful activity is used more at lower levels of output.

costly. Generalizing Polinsky and Shavell (1982) this paper explores several different functional forms of administrative costs in a setting with multiple externalities. Administrative costs may (1) be a function of tax rates because higher rates cause more evasion, (2) be a function of activity levels because higher levels of activity require more monitoring, (3) have fixed costs, or (4) be a any combination of (1) - (3).

If administrative costs are a function of taxes rates, the social planner should tax every activity. However, the administrative cost imposes a tradeoff on the policymaker. Higher tax rates reduce the externality but increase the administrative cost. In the single activity case, administrative costs that increase with the tax rate always lower the tax relative to the first-best case, which means that the externality is not fully corrected. In the multiple activity case, administrative costs that increase with tax rates also optimally leave externalities only partially corrected. If an activity has a large externality and its tax has a large administrative cost, it may be optimal to subsidize a less harmful substitute.

If administrative costs are a function of activity levels, it may no longer be optimal to tax every activity because the reduced external harm may be smaller than the administrative cost and lost private benefit. When it is optimal to tax every activity, the social planner should set the tax equal to the externality added to the marginal cost. At that tax rate, the private market internalizes both the externality and the administrative cost. When it is not optimal to tax every activity, the social planner must optimize under incomplete taxation. The analysis of fixed costs follows a similar line of reasoning. Lastly, section three explores how Pigouvian taxation in the second-best must be modified to account for a revenue requirement.

1.1 First-best Pigouvian taxation

This section introduces a model of Pigouvian taxation with multiple externalities and explores optimal policy when the social planner faces neither constraint nor cost. The result is as expected: the optimal tax vector is equal to the optimal vector of externalities.

Let x be an n-dimensional vector of activities.[5] Activities are used to generate goods and thus indirectly increase utility. Activities also have private costs and can only be performed in nonnegative quantities. Let $X \subseteq \mathbb{R}^n_+$ be the plausible set of activities.[6]

Following Ramsey (1927), a net benefit function $b : X \to \mathbb{R}$ maps activity levels to private net benefit. b is twice continuously differentiable, strictly concave, and achieves its

[5] We find activity levels the most natural interpretation. However, x could alternatively be interpreted as emission levels, consumption goods, or inputs. One disadvantage of interpreting x as a consumption good is that goods may be consumed in different ways that don't all cause the same external harm (Sandmo, 1978).

[6] This set is convex, open, has finite measure, and lies entirely in the positive orthant.

maximum somewhere in the interior of X.[7] If x consists of all productive activities, this is a general equilibrium model. As shown in the appendix, this net benefit function generalizes the maximization of a strictly increasing, strictly concave utility function subject to a convex production possibility frontier.

Both external harm and tax burden are linear functions of activities.[8] Let e be the n-dimensional vector of activity externalities and t be the n-dimensional vector of activity taxes. Tax revenues are assumed to be lump-sum redistributed.

Proposition 1.1.1. *In the first-best (i.e. when activity taxes are complete and tax administration is costless), the optimal tax vector is equal to the externality vector.*

Proof. The private market[9] solves the following problem:

$$\max_x b(x) - t^\top x$$

which leads to the first order condition $b'(x) - t^\top = 0$.[10] Because b is strictly concave it has an invertible Hessian. Therefore, by the implicit mapping theorem, there exists a continuously differentiable function, $x(t) : \mathbb{R}^n \to \mathbb{R}^n$, such that $b'(x(t)) - t^\top = 0$. $x(t)$ is the private market's best response function to the tax vector. Note that $b''(x(t))x'(t) = I$, the identity matrix, so $x'(t)$ is also invertible and $x'(t)^{-1} = t'(x)$. The social planner solves the following problem:

$$\max_t b(x(t)) - e^\top x(t)$$

which leads to the first order condition $b'(x(t^*))x'(t^*) - e^\top x'(t^*) = 0$. Substituting $b'(x(t^*)) = t^{*\top}$ yields $(t^* - e)^\top x'(t^*) = 0$. $t^* = e$ is clearly a solution, and the invertibility of $x'(t^*)$ ensures that it is the unique solution.[11] □

This is Pigou (1920)'s remarkable result generalized to arbitrary dimensions. When $t = e$, the private market fully internalizes every externality, and the policymaker does not need to

[7] One way to ensure this outcome is to assume that $\lim_{x_i \searrow \partial X} \frac{\partial b}{\partial x_i} = \infty$ and $\lim_{x_i \nearrow \partial X} \frac{\partial b}{\partial x_i} = -\infty$.

[8] Linearity makes the problem more tractable, but the main results would be preserved under weak convexity

[9] We describe the private market as an agent for brevity's sake. More precisely, individuals and firms make choices in the private market that result in an aggregate quantity of activity. We represent their behavior as a maximization problem. The solution to the maximization problem is the market equilibrium quantity of activity.

[10] We use matrix calculus and the associated notation.

[11] In the appendix we show that if there is no administrative cost t^* exists and is the unique global maximizer. Depending on the constraints imposed on the planner and the exact functional form of administrative cost it is possible that there exists no finite t^* or that t^* is not unique in the second-best.

know or use information other than the externality vector. Any other tax policy is strictly worse.[12] Using the tax on activity A to induce changes in activity B is not welfare improving because there is no benefit to changing the activity level in a market that already internalizes the externality.

1.2 Constrained Pigouvian taxation

This section modifies the model presented in the previous section by exploring two different constraints. First, the policymaker is constrained to tax only a subset of activities. We call this *incomplete taxation*. Second, the policymaker is constrained to tax only market transactions. We call this *output taxtion* in contrast to *activity taxation*. A policymaker may not be able to tax some activities because she faces political constraints or because the measurement of some activities is technologically impossible, prohibitively expensive, or particularly susceptible to evasion. Even if no activities can be taxed, the policymaker may be able to tax output, which is easier to measure because of the record keeping associated with market transactions.

1.2.1 Incomplete taxation

Leaving even one of the relevant activities untaxed alters the analysis because the tax vector no longer sets the marginal social benefit of each activity equal to the marginal social cost of each activity.

Example 1.2.1. *Let x_1 be operating a natural gas plant and x_2 be operating a coal plant to produce electricity. Assume there is only one tax, t_1 on x_1. The private market maximizes $b(x_1, x_2) - t_1 x_1$, leading to the first order conditions $\frac{\partial b}{\partial x_1} = t_1$ and $\frac{\partial b}{\partial x_2} = 0$. The social planner's problem is:*

$$\max_{t_1} b(x_1(t_1), x_2(t_1)) - e_1 x_1(t_1) - e_2 x_2(t_1)$$

with first order condition $\frac{\partial b}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial b}{\partial x_2}\frac{\partial x_2}{\partial t_1} - e_1\frac{\partial x_1}{\partial t_1} - e_2\frac{\partial x_2}{\partial t_1} = 0$. Substituting the private market first order condition leads to $(t_1^* - e_1)\frac{\partial x_1}{\partial t_1} - e_2\frac{\partial x_2}{\partial t_1} = 0$. Since $\frac{\partial x_1}{\partial t_1} < 0$ the optimal tax is:

$$t_1^* = e_1 + e_2\frac{\partial x_2}{\partial t_1}\frac{\partial t_1}{\partial x_1}$$

[12] This is true because the net benefit function is strictly concave. If activities were perfect complements or substitutes, the optimal tax vector would not be unique.

The sign of $\frac{\partial x_2}{\partial t_1}$ *is ambiguous and depends on the complementarity of* x_1 *and* x_2*. If the two activities are substitutes then* $\frac{\partial x_2}{\partial t_1} > 0$*; if they are complements then* $\frac{\partial x_2}{\partial t_1} < 0$*. A subsidy will be optimal if the taxed activity has the smaller externality and the two activities are very substitutable. In our example, if coal has a much larger external harm, natural gas is a substitute, and the social planner cannot administer a tax on coal, then a subsidy on natural gas will be optimal, even though using natural gas to generate electricity has an external harm.*

In the above example, the optimal tax depends on complementarity because the marginal private benefit of each untaxed activity will be 0 regardless of that activity's external harm. Because of this uncorrected externality the marginal social benefit of an untaxed activity will not equal that activity's marginal social cost. Using taxes to change the levels of untaxed, externally harmful activities improves welfare. At the optimum, the marginal social benefit of each tax is equal to that tax's marginal social cost. The marginal social benefit of a tax is the reduction in external harm of the taxed activity and all untaxed activities, and the marginal social cost of the tax is the lost private benefit. The following proposition generalizes the example to arbitrary dimensions.

Proposition 1.2.1. *If tax administration is costless, the optimal tax on each taxed activity is equal to the externality generated by that activity plus the externalities of all untaxed activities weighted by the responsiveness of the untaxed activity to changes in the taxed activity.*

Proof. Let Θ be the power set of $\{1, ..., n\}$, θ an arbitrary element of Θ, m the dimension of θ, x_θ the m-dimensional vector of taxed activities, $\bar{x}_\theta$ the $n-m$-dimensional vector of untaxed activities, t_θ the m-dimensional vector of taxes on x_θ, e_θ the m-dimensional vector of externalities generated by x_θ, and $\bar{e}_\theta$ the $n-m$-dimensional vector of externalities generated by $\bar{x}_\theta$. The private market solves the following problem:

$$\max_x b(x) - t_\theta^\top x_\theta$$

with solution $\frac{\partial b}{\partial x_j} = 0$ for $j \notin \theta$ and $\frac{\partial b}{\partial x_i} = t_i$ for $i \in \theta$. As before, the concavity of b ensures that a continuously differentiable best response function, $x(t_\theta)$ exists. The social planner solves the following problem:

$$\max_{t_\theta} b(x(t_\theta)) - e_\theta^\top x_\theta(t_\theta) - \bar{e}_\theta^\top \bar{x}_\theta(t_\theta)$$

which leads to the first order condition $b'(x(t_\theta^*))x'(t_\theta^*) - e_\theta^\top x_\theta'(t_\theta^*) - \bar{e}_\theta^\top \bar{x}_\theta'(t_\theta^*) = 0$. Note that the marginal harm of the tax is equal to the marginal benefit of the tax. Substituting the private market first order condition yields $(t_\theta^* - e_\theta)^\top x'_\theta(t_\theta^*) - \bar{e}_\theta^\top \bar{x}'_\theta(t_\theta^*) = 0$. Rearranging and applying the invertibilty of $x'_\theta(t_\theta^*)$ yields $t_\theta^{*\top} = e_\theta^\top + \bar{e}_\theta^\top \bar{x}'_\theta(t_\theta^*) x'_\theta(t_\theta^*)^{-1}$. For each individual tax:

$$t_i^* = e_i + \sum_{k \in \theta} \sum_{j \notin \theta} e_j \frac{\partial x_j}{\partial t_k} \frac{\partial t_k}{\partial x_i} \tag{1.1}$$

□

Note the special case, $m = n$ in which $x'(t_\theta^*) = x'_\theta(t_\theta^*)$ gives $t_\theta^* = e$ as before. If $m < n$ each activity tax may be above or below the associated externality; it may even have the opposite sign as the associated externality.

1.2.2 Output taxation

An output tax cannot generally induce the private market to select the socially optimal combination of activities. Whereas activity taxes can induce firms to substitute one activity for another, an output tax cannot. Consider, for example, producing electricity from either coal or natural gas. Assume using coal is privately cheaper (by an arbitrarily small amount) but also has a larger externality. An output tax will not discourage coal use relative to natural gas use, but an activity tax on coal will. In this particular case, since coal is only barely cheaper than natural gas and since they are perfect substitutes, an infinite tax (or any tax that makes natural gas privately preferable to coal) on coal will be optimal. Output taxes can cause the private market to substitute between activities if different combinations of activities are optimal at different scales of production. A tax on electricity would cause substitution from coal to natural gas if using coal exhibits better economies of scale.

The model used in the previous section is also used in this section albeit with some modification. In the previous section, the model made no explicit reference to output. In fact, the model could implicitly include many different outputs. In this section, we restrict the model to one output and introduce the increasing and strictly concave function $q(x) : X \to \mathbb{R}$ which maps the vector of activities to the quantity of output produced.[13]

If there is only one production activity and one output, the social planner can achieve the same welfare with either a tax on the activity or a tax on output. With output tax τ, the private market solves:

$$\max_x b(x) - \tau q(x)$$

[13] The appendix contains further discussion of $q(x)$.

which leads to the first order condition $b'(x) - \tau q'(x) = 0$, which defines a best response function $x(\tau)$.[14] The social planner solves the following problem:

$$\max_{\tau} b(x(\tau)) - ex(\tau)$$

which leads to the first order condition $b'(x(\tau^*))x'(\tau^*) - ex'(\tau^*) = 0$, which sets the marginal benefit of the tax equal to the marginal cost of the tax —where the benefit is reduced external harm and the cost is reduced private net benefit. Substituting the private market's first order condition, $b'(x(\tau^*)) = \tau^* q'(x(\tau^*))$, dividing by $x'(\tau^*)$, and rearranging yields:

$$\tau^* q'(x(\tau^*)) = e \implies \tau^* = e/q'(x(\tau^*)) \tag{1.2}$$

The equation on the left shows that at the optimum the marginal tax burden is equal to the marginal external harm. The marginal product of the activity scales the tax to properly align private incentives. Accordingly, when there is a single activity, the optimal activity tax and the optimal output tax yield the same welfare.

The remainder of this section considers taxing a single output when there are multiple activities.

Proposition 1.2.2. *At the optimal output tax, the marginal tax burden is equal to the marginal external harm of output.*

Proof. Let τ be the tax on output. Then the private market's problem is:

$$\max_{x} b(x) - \tau q(x)$$

with first order condition $b'(x(\tau)) - \tau q'(x(\tau)) = 0$, where $x(\tau) : \mathbb{R} \to \mathbb{R}^n$ is the private market's best response function to τ. Note that the effective marginal tax on an activity is simply the tax rate times the marginal product of each activity. The planner's problem is then:

$$\max_{\tau} b(x(\tau)) - e^\top x(\tau)$$

with first order condition $b'(x(\tau^*))x'(\tau^*) - e^\top x'(\tau^*) = 0$. Substituting in the private market optimum:

$$\tau^* q'(x(\tau^*))x'(\tau^*) = e^\top x'(\tau^*)$$

[14] The appendix contains a proof that $x(\tau)$ exists in the multidimensional case and also describes some of $x(\tau)$'s properties.

□

The optimal tax is a ratio of marginal harm to marginal product:

$$\tau^* = \frac{e^\top x'(\tau^*)}{q'(x(\tau^*))x'(\tau^*)} = \frac{e^\top A q'(x(\tau^*))^\top}{q'(x(\tau^*))Aq'(x(\tau^*))^\top} \tag{1.3}$$

where A is a negative definite matrix described in the appendix. The denominator is the marginal output of an increase in the output tax and is always negative. The numerator is the marginal external harm of an increase in the output tax and may be either positive or negative. The output tax will be larger when the activities that decrease the most in response to the output tax have large externalities. The output tax will have smaller magnitude if the activities that are most responsive to the output tax have large marginal products. The output tax will be negative if very harmful activities are relatively less productive at high levels of output.

Example 1.2.2. *Let x_1 be operating a natural gas plant and x_2 be operating a coal plant to produce electricity. Assume there is a tax τ levied on electricity. The private market maximizes $b(x_1, x_2) - \tau q(x_1, x_2)$ with first order conditions $\frac{\partial b}{\partial x_1} = \tau \frac{\partial q}{\partial x_1}$ and $\frac{\partial b}{\partial x_2} = \tau \frac{\partial q}{\partial x_2}$. The social planner's problem is:*

$$\max_{\tau} b(x_1(\tau), x_2(\tau)) - e_1 x_1(\tau) - e_2 x_2(\tau)$$

with first order condition $\frac{\partial b}{\partial x_1}\frac{\partial x_1}{\partial \tau} + \frac{\partial b}{\partial x_2}\frac{\partial x_2}{\partial \tau} - e_1\frac{\partial x_1}{\partial \tau} - e_2\frac{\partial x_2}{\partial \tau} = 0$. Substituting in the private market optimum $\tau^*\frac{\partial q}{\partial x_1}\frac{\partial x_1}{\partial \tau} + \tau^*\frac{\partial q}{\partial x_2}\frac{\partial x_2}{\partial \tau} - e_1\frac{\partial x_1}{\partial \tau} - e_2\frac{\partial x_2}{\partial \tau} = 0$ The optimal tax is:

$$\tau^* = \frac{e_1\frac{\partial x_1}{\partial \tau} + e_2\frac{\partial x_2}{\partial \tau}}{\frac{\partial q}{\partial x_1}\frac{\partial x_1}{\partial \tau} + \frac{\partial q}{\partial x_2}\frac{\partial x_2}{\partial \tau}}$$

By assumption both coal plants and natural gas plants pollute ($e_1 > 0$ and $e_1 > 0$) and more electricity is produced if plants increase operations ($\frac{\partial q}{\partial x_1} > 0$ and $\frac{\partial q}{\partial x_1} > 0$). The signs of $\frac{\partial x_1}{\partial \tau}$ and $\frac{\partial x_1}{\partial \tau}$ are ambiguous, but at least one of them must be negative. A subsidy will be optimal if the use of coal increases with the tax ($\frac{\partial x_2}{\partial \tau} > 0$), coal produces a large externality relative to natural gas ($e_2 > e_1$), and the marginal product of natural gas is large relative to the marginal product of coal ($\frac{\partial q}{\partial x_1} > \frac{\partial q}{\partial x_2}$).

In order to make a welfare comparison between the optimal output tax and the optimal activity tax, we find a map from the output tax to activity taxes.

Proposition 1.2.3. *For any output tax the private market will respond as if there is a tax on each activity equal to the output tax times the marginal product of that activity.*

Proof. We want to find $t(\tau)$ such that $x(t(\tau)) = x(\tau)$. Recall that $b'(x(t)) = t^\top$ and $b'(x(\tau)) - \tau q'(x(\tau)) = 0$. Thus

$$t(\tau)^\top = \tau q'(x(\tau)) \tag{1.4}$$

□

Substituting the optimal output tax derived above:

$$t(\tau^*)^\top = \frac{e^\top x'(\tau^*)}{q'(x(\tau^*))x'(\tau^*)} q'(x(\tau^*)) \tag{1.5}$$

The higher the marginal product of an activity, the higher the effective tax on that activity. This relationship makes sense because with a tax on output, an increase in an activity causes an increase in the tax burden proportional to the activity's marginal product.

In the appendix we show how welfare changes when moving from one activity tax to another. Applying that derivation shows that the welfare lost if the social planner is constrained to tax only output is:

$$\Delta b - e^\top \Delta x = (t(\tau^*) - e)^\top \left[\int_0^1 r \ x'\left(\gamma(r)\right) \ dr \right] (t(\tau^*) - e) \tag{1.6}$$

where $\gamma(r) = e + r(t(\tau^*) - e)$. Note that the integral is the sum of negative definite matrices and that the expression is a quadratic form. Thus the optimal activity tax is always weakly better than any output tax.[15] Note that the welfare lost increases with the square of the uncorrected externality and increases with the weighted average size of $x'(t)$ over the set $[e, t(\tau^*)]$. $x'(t)$ is weighed by r for the same reason dead weight loss approximations are a triangle; close to the optimum the marginal benefit and marginal social harm are equal but diverge the further $t(\tau^*)$ is from e.

In the appendix we also derive a Harberger (1964)-style approximation of the welfare change caused by moving from one activity tax vector to another, and we apply that method here to approximate the lost welfare caused by moving from the optimal activity tax to the

[15] If the optimal activity tax is equal to marginal product times the optimal output tax, then both taxes achieve the same welfare because $t(\tau^*) = e$.

optimal output tax.

$$\Delta b - e^\top \Delta x \approx \frac{1}{2}(t(\tau^*) - e)^\top (x(e) - x(\tau^*)) \tag{1.7}$$
$$= \frac{1}{2}(\tau^* q'(x(\tau^*)) - e)^\top (x(e) - x(\tau^*)) \tag{1.8}$$

At the optimal activity tax there are no uncorrected externalities. At the optimal output tax the uncorrected externality is $(t(\tau^*) - e)$ or $(\tau^* q'(x(\tau^*)) - e)$. Note that the uncorrected externality is small if activities with large external harms also have high marginal products because the output tax discourages activities with high marginal products relatively more. Using the approximation, the average uncorrected externality for each activity is half of the uncorrected externality under the output tax, $\frac{1}{2}(t(\tau^*) - e)$. Multiplying the uncorrected externality by the 'excess' amount of the activity that occurs under the output tax, $(x(e) - x(\tau^*))$ yields the total welfare loss.

1.3 Costly Pigouvian taxation

Administrative cost includes expenditures on measurement, enforcement, collections, legislation, and litigation. Some of these costs appear in the Internal Revenue Service budget, which was 11.5 billion dollars for fiscal year 2015.[16] We call the case when administrative cost increases with activity levels *measurement costs*. Measurement costs arise because it is costly to determine the level of pollution generating activities, regardless of who is making these measurements and even if all parties are behaving honestly. Specific examples of these costs include the monitoring devices and the scientists who design and operate them. We call the case when administrative cost increases with tax rates *enforcement costs*. Enforcement costs arise if evasion increases with tax rates and the government pours more resources into tax enforcement, including more auditors, more lawyers, and more evasion detection software. We call the case when administrative cost increases with tax revenue collected *bureaucracy costs*. Bureaucracy costs resemble the Flypaper effect (Hines and Thaler, 1995)—larger revenues induce larger bureaucracies—for whatever reason the money sticks.

1.3.1 Costly administration with one activity

This section describes how administrative cost changes the optimal tax when there is only one activity and that activity generates an externality.

The social planner must construct an administrative apparatus to collect and enforce

[16] See Internal Revenue Service (2016) for additional information on the IRS budget.

taxes. Let c map the activity level and tax rate to administrative cost. Formally c is continuously differentiable and weakly convex and $\arg\min(c) = (0, 0)$. These assumptions are made for tractability, but they are consistent with a planner who employs the most effective tax collecting and enforcement resources first. Administrative cost and marginal administrative cost (with respect to x) are both increasing in activity level. A subsidy should also be costly to administer, so administrative cost and marginal administrative cost (with respect to t) are both increasing in the tax rate above 0 and are both decreasing in the tax rate below 0.

The social planner solves the following problem:

$$\max_t b(x(t)) - ex(t) - c(x(t), t)$$

which leads to the first order condition $b'(x(t^*))\frac{\partial x}{\partial t} - e\frac{\partial x}{\partial t} - c_1\frac{\partial x}{\partial t} - c_2 = 0$ where c_i denotes the partial derivative of c with respect to its i^{th} argument. Note the general rule: the marginal social benefit of the tax (reduced externality and decreased administrative cost) is equal to the marginal social cost of the tax (reduced private benefit and increased administrative cost). Substituting $b'(x(t^*)) = t^*$, dividing by $\frac{\partial x}{\partial t}$, and rearranging yields:

$$t^* = e + c_1 + c_2\frac{\partial t}{\partial x} \tag{1.9}$$

The following table presents the optimal tax associated with several possible functions of administrative cost. Note that for all cases with administrative cost the optimal tax expressions are implicit formulas.

Table 1.2 Single activity optimal taxes

Case	Cost function	Optimal Pigouvian tax	Previous Reseach
No cost	0	$t^* = e$	Pigou (1920)
Enforcement costs	$c(t)$	$t^* = e + \frac{\partial c}{\partial t}\frac{\partial t}{\partial x}$	
Measurement costs	$c(x)$	$t^* = e + \frac{\partial c}{\partial x}$	Polinsky & Shavell (1982)
Bureaucracy costs	$c(xt) = c(R)$	$t^* = e + \frac{\partial c}{\partial R}\left(t^* + x(t^*)\frac{\partial t}{\partial x}\right)$	Polinsky & Shavell (1982)
Arbitrary costs	$c(x, t)$	$t^* = e + c_1 + c_2\frac{\partial t}{\partial x}$	

When administrative cost takes the $c(t)$ functional form, the marginal administrative cost at $t = 0$ is 0. Thus the policymaker will always implement a tax. However, the tax is not equal to the externality and at the optimum the marginal social benefit of the activity is smaller than the marginal social cost of the activity. This is because when administrative cost is a function of tax rates, higher taxes reduce the external harm but increase administrative cost. The marginal benefit of a higher tax is lower external harm, and the marginal cost of a higher tax is lower private net benefit and higher administrative cost. Thus the tax rate is always lower than the case with no administrative cost,[17] which leaves an uncorrected externality. This shows up in the model because $\frac{\partial c}{\partial t}/\frac{\partial x}{\partial t} < 0$.[18]

When administrative cost takes the $c(x)$ functional form, there is a discrete jump in the administrative cost if policy maker increases the tax from 0. It is possible that $b(x(0)) - ex(0) > b(x(t^*)) - ex(t^*) - c(x(t^*))$ in which case the optimal tax is $t = 0$, and the externality is uncorrected. When implementing a tax is optimal, the optimal tax is not equal to the externality. If administrative cost is a function of x, the social planner should raise the tax until *the marginal social benefit of the activity equals the marginal social cost of the activity*. Because administrative cost increases only with x the administrative cost can be interpreted as an additional externality. Thus the optimal tax induces the private market to fully internalize the externality and the administrative cost. The tax rate is always higher than the case with no administrative cost.[19] This shows up in the model because $\frac{\partial c}{\partial x} > 0$.

1.3.2 Variable administrative costs

As in the one-dimensional model, when administrative cost is a function of activity levels, taxes should equal the externality plus the marginal administrative cost, and taxes will be higher compared to the costless case. At the optimal tax the private market fully internalizes the external harm and administrative cost. Thus the optimal tax expression does not include cross tax effects.

When administrative cost is a function of tax rates, taxes cannot cause the private market to fully internalize the externality. The planner should use taxes with low marginal administrative cost to affect other activity levels. Thus the optimal tax expression does include cross tax effects. The social planner should set the marginal benefit of each tax equal to the marginal cost of that tax.

The cost function $c(x, t)$, $c : X \times \mathbb{R}^n \to \mathbb{R}$, is the multidimensional analogue of the single activity case. We assume that $c(x, t)$ is continuously differentiable, weakly convex, and has $\arg\min(c) = (0, 0)$.

Example 1.3.1. *Let x_1 be operating a natural gas plant and x_2 be operating a coal plant to produce electricity. Assume that both activities are taxed and administrative cost increases*

[17] Similarly, if there is a positive externality the subsidy rate is always lower than the case with no administrative cost.

[18] For positive externalities $\frac{\partial c}{\partial t}/\frac{\partial x}{\partial t} > 0$. In either case the optimal tax will never be 0 because the marginal administrative cost is 0 at $t = 0$.

[19] However, the optimal subsidy is always smaller than the case with no administrative cost.

with t. The private market maximizes $b(x_1, x_2) - t_1 x_1 - t_2 x_2$, leading to the first order conditions $\frac{\partial b}{\partial x_1} = t_1$ and $\frac{\partial b}{\partial x_2} = t_2$. The concavity of b ensures that $x_1(t_1, t_2)$ and $x_2(t_1, t_2)$ exist. The social planner's problem is:

$$\max_t b(x_1(t_1, t_2), x_2(t_1, t_2)) - e_1 x_1(t_1, t_2) - e_2 x_2(t_1, t_2) - c(t_1, t_2)$$

with first order conditions:

$$\begin{aligned}
\frac{\partial b}{\partial x_1}\frac{\partial x_1}{\partial t_1} + \frac{\partial b}{\partial x_2}\frac{\partial x_2}{\partial t_1} - e_1\frac{\partial x_1}{\partial t_1} - e_2\frac{\partial x_2}{\partial t_1} - \frac{\partial c}{\partial t_1} = 0 \ \textit{and} \\
\frac{\partial b}{\partial x_1}\frac{\partial x_1}{\partial t_2} + \frac{\partial b}{\partial x_2}\frac{\partial x_2}{\partial t_2} - e_1\frac{\partial x_1}{\partial t_2} - e_2\frac{\partial x_2}{\partial t_2} - \frac{\partial c}{\partial t_2} = 0
\end{aligned}$$

Substituting the private market first order condition and manipulating the equation leads to:

$$\begin{aligned}
t_1^* = e_1 + \frac{\partial c}{\partial t_1}\frac{\frac{\partial x_2}{\partial t_2}}{\frac{\partial x_1}{\partial t_1}\frac{\partial x_2}{\partial t_2} - \frac{\partial x_1}{\partial t_2}\frac{\partial x_2}{\partial t_1}} - \frac{\partial c}{\partial t_2}\frac{\frac{\partial x_2}{\partial t_1}}{\frac{\partial x_1}{\partial t_1}\frac{\partial x_2}{\partial t_2} - \frac{\partial x_1}{\partial t_2}\frac{\partial x_2}{\partial t_1}} \ \textit{and} \\
t_2^* = e_2 + \frac{\partial c}{\partial t_2}\frac{\frac{\partial x_1}{\partial t_1}}{\frac{\partial x_1}{\partial t_1}\frac{\partial x_2}{\partial t_2} - \frac{\partial x_1}{\partial t_2}\frac{\partial x_2}{\partial t_1}} - \frac{\partial c}{\partial t_1}\frac{\frac{\partial x_1}{\partial t_2}}{\frac{\partial x_1}{\partial t_1}\frac{\partial x_2}{\partial t_2} - \frac{\partial x_1}{\partial t_2}\frac{\partial x_2}{\partial t_1}}
\end{aligned}$$

Noting that $x'(t)^{-1} = t'(x)$, we have:

$$\begin{aligned}
t_1^* = e_1 + \frac{\partial c}{\partial t_1}\frac{\partial t_1}{\partial x_1} + \frac{\partial c}{\partial t_2}\frac{\partial t_2}{\partial x_1} \ \textit{and} \\
t_2^* = e_2 + \frac{\partial c}{\partial t_2}\frac{\partial t_2}{\partial x_2} + \frac{\partial c}{\partial t_1}\frac{\partial t_1}{\partial x_2}
\end{aligned}$$

A subsidy on natural gas will be optimal if the cost of administering the subsidy is small ($\frac{\partial c}{\partial t_1}$ is negative and has a small magnitude), administering the tax on coal is costly ($\frac{\partial c}{\partial t_2}$ is positive and has a large magnitude), coal is more harmful than natural gas ($e_2 > e_1$), and coal x_2 and natural gas x_1 are very substitutable ($\frac{\partial x_2}{\partial t_1}$ is positive with large magnitude).

In arbitrary dimensions, the private market problem is the same as in the multiple activity, complete taxation, costless administration case. The social planner solves:

$$\max_t b(x(t)) - e^\top x(t) - c(x(t), t)$$

which leads to the first order condition $b'(x(t^*))x'(t^*) - e^\top x'(t^*) - c_1(x(t^*), t^*)x'(t^*) - c_2(x(t^*), t^*) = 0$, where c_1 denotes the partial derivative of c with respect to its first vector argument and c_2 is the partial derivative of c with respect to its second vector argument. Substituting in the private market first order condition and applying $x'(t^*)$'s invertibility yields:

$$t^{*\top} = e^\top + c_1 + c_2 x'(t^*)^{-1} \tag{1.10}$$

The table below describes some special cases. Note that for both cases with administrative cost the optimal tax expressions are implicit formulas.

Table 1.3 Multiple activity optimal taxes

Case	Cost function	Optimal Pigouvian tax matrix notation	Optimal Pigouvian tax element notation
No administrative cost	$c = 0$	$t^{*\top} = e^\top$	$t_i^* = e_i$
Measuring costs	$c = c(x)$	$t^{*\top} = e^\top + c'(x(t^*))$	$t_i^* = e_i + \frac{\partial c}{\partial x_i}$
Enforcement costs	$c = c(t)$	$t^{*\top} = e^\top + c'(t^*)x'(t^*)^{-1}$	$t_i^* = e_i + \sum_j \frac{\partial c}{\partial t_j}\frac{\partial t_j}{\partial x_i}$

Just as in the single activity case, if administrative cost is a function of only x, the administrative cost can be fully internalized.[20]

Proposition 1.3.1. *If administration is costly and the cost is a function of activity levels, the optimal tax vector is equal to the externality vector added to the marginal administrative cost.*

If administrative cost is a function of x, the social planner should raise the tax until *the marginal social benefit of the activity equals the marginal social cost of the activity.* The administrative cost can be interpreted as an additional externality. The characterization of the optimum provides an implicit solution—both sides of the equation are functions of t^*. Nonetheless there is no complementarity: $x'(t^*)$ does not enter into the equation. This is not the case when administration is a function of t.

Proposition 1.3.2. *If administration is costly and the cost is a function of the tax rates, the optimal tax on each activity is equal to the externality generated by that activity plus the sum of all the marginal administrative costs weighted by the responsiveness of the tax rate to changes in activity.*

[20] The administrative cost function could change if some taxes are not used. As with incomplete taxation, there is no analytical solution—the social planner must find the optimal tax for each set of taxes and chose the one that yields the highest welfare. There are up to 2^n different possible administrative cost functions, reflecting the monitoring costs for every possible set of activity taxes. Proposition 3 assumes that the cost function does not change with the tax base.

If administrative cost varies with t, the optimal tax does not exhibit complementarity irrelevance. The social planner should raise taxes until *the marginal social benefit of the tax is equal to the tax's marginal social harm.* A subsidy may be optimal for some externally harmful activities.[21] c' will take negative values whenever there is a subsidy because the cost of administration will decrease the less negative the tax becomes.

1.3.3 Fixed administrative costs

If there are fixed administrative costs, it may not be optimal to tax some activities at all—leading to incomplete taxation. Leaving even one of the relevant activities untaxed alters the analysis (compared to complete taxation) because the tax vector no longer sets the marginal benefit of each activity equal to the marginal harm of each activity—for all untaxed activities, the marginal private benefit is equal to 0 regardless of the tax on the other variables. At the optimum each tax must not only account for the activity it is applied to but also every other untaxed activity.

If each activity tax has its own fixed cost, the social planner must optimize the social welfare function 2^n times—once for each possible combination of taxes—using incomplete taxes as described above. Let the fixed cost for each tax be f_i. Recall that the power set of $\{0, ..., n\}$ is Θ. The social planner's problem is:

$$\max_{\theta \in \Theta} \left\{ \max_{t_\theta} b(x(t_\theta)) - e^\top x(t_\theta) - \sum_{i \in \theta} f_i \right\}$$

No closed form solution exists, but the optimal tax expression for each subproblem is:

$$t_\theta^{*\top} = e_\theta^\top + \bar{e}_\theta^\top \bar{x}'_\theta(t_\theta^*) x'_\theta(t_\theta^*)^{-1} \tag{1.11}$$

Variable administrative cost $c(x_\theta(t_\theta), t_\theta)$ may be included in which case the optimal tax expression for each subproblem is:

$$t_\theta^{*\top} = e^\top x'(t_\theta^*) x'_\theta(t_\theta^*)^{-1} + c_1 + c_2 x'_\theta(t_\theta^*)^{-1} \tag{1.12}$$

The same intuition discussed above applies in this case.

[21] The matrix $t'(x(t^*)) = b''(x(t^*))$ describes the effect of a change in the activity vector on tax rates at the optimal tax rate. Because b is concave, this matrix is negative definite, so the diagonals are all negative. An increase in t_i will, thus, reduce x_i although it may increase or have no effect on x_j. This implies that t^* may have negative entries.

1.3.4 Variable administrative cost output tax

Output taxes with variable administrative cost are similar to a single activity tax with variable administrative cost. If administrative cost is a function of output level, the output tax will be higher compared to the costless case. If administrative cost is a function of the output tax rate the optimal output tax will be lower than the costless case.

Let $d : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ map quantity and output tax rate to administrative cost. Formally d is continuously differentiable, weakly convex, and has $\arg\min(d) = (0, 0)$. The private market problem is the same as in the costless output tax case. The social planner solves:

$$\max_{\tau} b(x(\tau)) - e^\top x(\tau) - d(q(x(\tau)), \tau)$$

with first order condition:

$$b'(x(\tau^*))x'(\tau^*) - e^\top x'(\tau^*) - d_1 q'(x(\tau^*))x'(\tau^*) - d_2 = 0$$

where d_i represents the derivative of d with respect to its i^{th} argument. Substituting the private market first order condition $b'(x(\tau^*)) = \tau^* q'(x(\tau^*))$ yields:

$$\tau^* = \frac{e^\top x'(\tau^*) + d_1 q'(x(\tau^*))x'(\tau^*) + d_2}{q'(x(\tau^*))x'(\tau^*)} \tag{1.13}$$

$$= \frac{e^\top x'(\tau^*)}{q'(x(\tau^*))x'(\tau^*)} + d_1 + \frac{d_2}{q'(x(\tau^*))x'(\tau^*)} \tag{1.14}$$

If administrative cost varies with output, the optimal output tax will be larger. If administrative cost varies with the output tax rate, the optimal output tax will be smaller. To see this note that $d_1 > 0$, $d_2 > 0$, and $q'x' < 0$.

If the problem includes both activity and output taxes the private market solves:

$$\max_{x} b(x) - t^\top x - \tau q(x)$$

With the most general form of variable administrative cost, the social planner solves:

$$\max_{t,\tau} b(x(t,\tau)) - e^\top x(t,\tau) - c(x(t,\tau), t) - d(q(x(t,\tau)), \tau)$$

However, the private market's problem does not yield a one-to-one best response function, $x(t, \tau)$—there are infinitely many t's and τ's that lead to the same x. Because of this the implicit expressions that come from the first order conditions of the social planner's problem are only informative about t and τ jointly. The end result is that little analysis can be performed.

1.3.5 Revenue requirement

Pigouvian taxes generate revenue in addition to aligning incentives. In this subsection we show how the optimal taxes described above are altered when there is a revenue constraint. Assume that the revenue requirement is $R \geq t_\theta^\top x_\theta(t_\theta)$. We begin with the Lagrangian:

$$\mathcal{L} = b(x(t_\theta)) + \lambda(R - t_\theta^\top x_\theta(t_\theta))$$

with first order condition:

$$b'(x(t_\theta^*))x'(t_\theta^*) - \lambda(t_\theta^{*\top} x_\theta'(t_\theta^*) + x_\theta(t_\theta^*)^\top) = 0$$

which sets the marginal social cost of each tax equal to the marginal revenue of that tax times the Lagrange multiplier and leads to the optimal tax expression:

$$t_\theta^{*\top} = \frac{\lambda}{1-\lambda} x_\theta(t_\theta^*)^\top x_\theta'(t_\theta^*)^{-1} \tag{1.15}$$

Note that the Karush Kuhn Tucker conditions require that $\lambda \leq 0$, which makes sense since increasing the required revenue should decrease private net benefit. Now we determine the optimal tax when there is a revenue constraint and Pigouvian taxation using the most general model with incomplete taxation and costly administration.

Proposition 1.3.3. *The optimal Pigouvian tax with a revenue constraint is equal to the expression for the optimal revenue constrained tax added to the expression for the optimal unconstrained Pigouvian tax times a scaling factor.*

Proof. Starting with the Lagrangian:

$$\mathcal{L} = b(x(t_\theta)) - e^\top x(t_\theta) - c(x_\theta(t_\theta), t_\theta) + \lambda(R - t_\theta^\top x_\theta(t_\theta))$$

with first order condition:

$$b'(x(t_\theta^*))x'(t_\theta^*) - e^\top x'(t_\theta^*) - c_1 x_\theta'(t_\theta^*) - c_2 - \lambda t_\theta^{*\top} x_\theta'(t_\theta^*) - \lambda x_\theta(t_\theta^*)^\top = 0$$

which sets the marginal social cost of each tax equal to the marginal revenue of that tax times the Lagrange multiplier. Rearranging:

$$(1-\lambda)t_\theta^{*\top} x_\theta'(t_\theta^*) = e^\top x'(t_\theta^*) + c_1 x_\theta'(t_\theta^*) + c_2 + \lambda x_\theta(t_\theta^*)^\top$$

The optimal tax expression is:

$$t_\theta^{*\top} = \frac{1}{1-\lambda} \underbrace{\left(e^\top x'(t_\theta^*) x'_\theta(t_\theta^*)^{-1} + c_1 + c_2 x'_\theta(t_\theta^*)^{-1}\right)}_{\text{optimal Pigouvian tax}} + \underbrace{\frac{\lambda}{1-\lambda} x_\theta(t_\theta^*)^\top x'_\theta(t_\theta^*)^{-1}}_{\text{optimal revenue constrained tax}} \tag{1.16}$$

□

The Karush Kuhn Tucker conditions require that $\lambda \leq 0$. If the revenue generated by the Pigouvian tax exceeds R, then $\lambda = 0$ and the optimal tax is equal to the Pigouvian tax.

This version of the Pigouvian tax problem has been analyzed by many papers. Sandmo (1975) suggests an additivity property:[22]

> [T]he optimal tax structure is characterized by what might be called an additivity property; the marginal social damage of commodity m enters the tax formula for that commodity additively, and does not enter the tax formulas for the other commodities, regardless of the pattern of complementarity and substitutability.

Our result generalizes Sandmo (1975)'s additivity property. Strictly speaking, Sandmo's additivity does not hold in our more general setting because complementarity matters in the optimal tax formula.[23] However, the optimal tax is the expression for the optimal Pigouvian tax multiplied by $\frac{1}{1-\lambda}$ added to the expression for the optimal revenue constrained tax.

1.4 Conclusion

Pigou's seminal insight demonstrated that in a first-best world taxes can fully correct externalities. This paper extends his insight into a second-best world, in which the policymaker faces constraints and costs. In this world there are several cases in which an externality cannot or should not be fully corrected and even cases in which a harmful activity should be subsidized. Intuitively, the policymaker should use the cheapest tax instruments available to her that effect the greatest reduction in external harm.

[22] See Kopczuk (2003), generalizing Sandmo's additivity property; Bovenberg and De Mooij (1994), describing the relationship between a tax on labor and a Pigouvian tax on a polluting consumption good; and Kaplow (2012), demonstrating that changing a commodity tax to the first best Piouvian tax while making compensating changes in the income tax creates a Pareto improvement in a model where utility is separable in labor and other activities.

[23] The literature reinforcing complementarity irrelevance is large. See Bovenberg and De Mooij (1994), Bovenberg and van der Ploeg (1994), Bovenberg and Goulder (1996), Fullerton (1997), Pirttilä and Tuomala (1997), Ng (1980), Kopczuk (2003), and Kaplow (2012). A notable exception is Cremer, Gahvari and Ladoux (1998), where complements and substitutes matter in a model that includes heterogeneous consumers and nonlinear external harm.

We highlight energy production examples, but there are other policy relevant examples of harm subsidies, including needle exchanges. Needle exchanges subsidize the use of heroin with clean needles in an attempt to induce substitution away from the use of heroin with dirty needles. Both activities are thought to be socially harmful, but using dirty needles is more harmful. If second-best considerations prevent the full correction of both activities, it may be optimal to subsidize the clean needles.

Our paper also suggests that optimal policy requires more than determining marginal external harm. In the second-best, the net private benefit function and administrative cost function are also relevant for policy decisions. While the examples of subsidizing harm suggest that policymakers intuitively understand how to set optimal policy in the second-best, our hope is that this paper helps formalize that intuition. Ideally, policymakers will collect the relevant market and administrative cost data to be able to set more precise policy. Future work could estimate optimal policy parameters and identify other interesting optimal harm subsidies.

CHAPTER II

How to Catch Capone: The Optimal Punishment of Interrelated Crimes

From a work with Daniel Jaqua

Abstract

This paper characterizes optimal criminal punishments when there are multiple interrelated crimes. Optimal punishments are functions of the extent to which related crimes are complements or substitutes weighted by their relative harms to society. The available empirical evidence on the relationship between index crimes in the United States suggests that tailoring criminal punishments properly to incorporate relationships between crimes could reduce the aggregate harm to victims by 3%, or about $8 billion dollars annually, holding enforcement expenditures fixed. The actual harm reduction of a marginal increase in arrests for an index crime is on average about 1.5-3 times greater than the harm reduction calculated without these effects.

Keywords: Crime, Criminal sanctions, Complement and substitute crimes
JEL Codes: D11, K14

2.1 Introduction

Crime places a heavy burden on society. A partial list of costs includes the direct harm from crime, decreased property values, policing, arresting, trying, convicting, sentencing, and incarcerating criminals, the lost contributions of those incarcerated, and the onus of wrongful convictions. Governments in the United States spent more than $260 billion in 2010 to administer criminal justice,[1] and Anderson (1999) estimates the total burden of crime to be $1.7 trillion. Sentencing crimes optimally should decrease these burdens.

Most crimes are not committed in isolation. Al Capone headed a criminal organization that profited from bootlegging, prostitution, racketeering, and murder. He was tried and served a prison term, but his convictions were not for any of these crimes. He served time for tax evasion and contempt of court. In Capone's case, the rules underlying the criminal justice system were certainly stretched (and possibly broken) to ensure that he ended up behind bars. Nonetheless his experience illustrates that (1) criminals often undertake many criminal actions; (2) the returns to these actions are a function of other criminal acts; (3) these actions vary in their social harm; and (4) these actions also vary in how costly they are to detect and punish.[2] Thanks to the relationship between Capone's more socially harmful crimes and tax evasion, it was probably optimal to convict him for tax evasion and give him the most severe sentence possible. More generally, basing part of the enforcement of crimes on related criminal activities can reduce the social burden of criminal activities.

A criminal may commit several different crimes contemporaneously, the same crime serially, or different crimes non-contemporaneously. This is not random chance; in many cases criminals can increase the profit of a crime by committing other crimes in conjunction or in series. Previous work has touched on this, suggesting for example that economies of scale or learning might increase the payoff from serial repetition (Aizer and Doyle, 2015). While both of these models suggest specialization, surveys of prisoners and records of criminals re-arrested for different crimes demonstrate that many criminals are generalists (Beck and Shipley, 1989). Empirical work supports this claim, finding that changing the punishment for one crime may have large and significant effects on the commission of other crimes (Levitt, 1998*b*; Kuziemko and Levitt, 2004; Benson, Rasmussen and Kim, 1998; Shepherd, 2002).

If increasing the punishment of a crime leads to decreased commission of other crimes, then there is additional motivation to increase that punishment. If increasing the punishment of a crime leads to increased commission of other crimes, then there is cause to doubt that

[1] U.S. Bureau of Justice Statistics.

[2] A cost may be monetary but need not be. It may have been possible to convict Capone for his more socially harmful crimes, but doing so would have required an infringement on his rights that was socially too costly to bear.

the punishment should be increased. Taking these cross-crime effects into consideration could reduce the total burden of crime. Our estimates suggest that reallocating existing enforcement resources could reduce the harm to victims of index crimes by about $8 billion annually, which is about 3% of the total annual harm to victims of index crimes.[3]

Our paper analyzes interrelated crimes. However, our model's implications extend beyond criminal justice in two important ways. First, there is no need for the actions analyzed to be socially harmful acts. Second, there is no need for the punishment to be administered through the criminal justice system.

We illustrate the first point with the example of money laundering. In isolation, money laundering is not socially harmful–the launderer is simply obscuring her source of income. The benefit of obscuring the source of income is high when that income is criminally derived or deployed, and this is why money laundering is criminalized. Stated differently, the punishment of money laundering is aimed at decreasing the commission of other acts, not money laundering itself. Similar logic applies to many other activities including many possessory and conspiratorial crimes.

The second extension of the model is to sanctions outside of the criminal justice system. The two most obvious bodies under which sanctions are issued are civil law courts and government regulatory agencies. Tort law affects the propensity to commit litigable acts, but it surely encourages and discourages other acts too. Government regulation, for example through corrective taxation, can be analyzed in the same model. This approach has purchase even outside of government structures–parents may set curfews (and corresponding punishments) for several reasons and likely not only because they object to their children being out late. Curfews may increase children's sleep and homework completed or decrease their drug-use and hooliganism.

Legislatures, judiciaries, and prosecutors already informally take into consideration crime interdependence. Famously, Al Capone was convicted of tax evasion because enforcement agencies were unable to convict him for any of his other, more socially harmful, criminal acts. When statutory punishments are enacted, legislative records suggest that interrelation is sometimes taken into consideration. For example, legislators have made claims about the effect of illegal immigration on criminal activity in support of a particular sanction on illegal immigration, and similar arguments are made concerning illegal drug use. The legislators' analysis underlying these claims, however, may not be rigorous. We aim to provide machinery to improve this analysis.

[3] The FBI designates certain crimes as index crimes and uses these crimes to produce its annual report. These crimes have a standardized definition across states. In this paper, we consider the following index crimes: murder and non-negligent manslaughter; forcible rape; robbery, assault; burglary; larceny; and auto-theft.

In the following section we present a model of criminal activity and characterize the social planner's optimal policy response. Section 2 explains the difference between this result and other analyses of criminal activity and relates this result to an abstract Pigouvian tax setting. Section 3 relates insights about the optimal policy to observations of the criminal justice system. Section 4 discusses prior estimates of the own-price elasticity of crime and the few studies which attempt to measure some type of cross price elasticity and also converts existing estimates into a welfare measure of the relative importance of our paper's contribution.

2.2 A Model of Interrelated Crimes

Our model considers individuals who allocate their time to various activities under a temporal budget constraint.[4] These activities generate utility either directly or by generating income that can be used for consumption thus indirectly increasing utility. Some activities are socially harmful and some are socially harmless.[5] We call the socially harmful activities crimes. This is in contrast to the common definition of crime as an activity which the law criminalizes–a definition we cannot use since punishment is endogenously determined.

A social planner may choose to punish activities. A punishment is an additional cost, borne by the criminal, of committing a crime. An obvious intuitive interpretation of punishment is an additional expected time cost of committing a crime, equal to the probability of conviction multiplied by the length of prison sentence or community service for an activity. Formally, these are the punishments in the model, but the insights are valid for other sanctions. For example, rehabilitation, fines, or a job training program could each be prescribed responses to conviction. We model the behavior of the individual as responding to a time cost, but the model could be extended to explicitly analyze more complex inputs into the individual's decision making process.

2.2.1 Representative Agent Model

Becker (1968) models the decision to commit crime as the result of the maximization of a Von Neuman Morgenstern utility function $EU = pU(Y-f)+(1-p)U(Y)$.[6] Where Y is the wealth of the individual, inclusive of psychic gain, f is the fine, and p is the probability of punishment. Becker's contribution focuses on the insight that criminals respond to incentives

[4] Heineke (1978); Ehrlich (1973) also uses this approach to individual behavior

[5] A more general version of the model would allow socially beneficial activities and describe the optimal policies governing them. The intuition of such a model is the same, but some activities would be subsidized.

[6] In a footnote, Becker recognizes that the payoffs for a given crime may depend on other aspects of the criminal justice system, particularly the punishments for other crimes because of substitution, but his analysis considers only one crime.

in the form of the probability p and severity f of punishment.

According to Becker's famous result, high sanctions that completely deter are costless and optimal. As previous research has shown, this result does not hold when complete deterrence is infeasible. Since some crime will occur, high or maximal sanctions are not optimal because they must still be administered. This result also does not hold if threatening sanctions is costly. The costliness of threatened sanctions is a realistic assumption for a variety of reasons, the most obvious being that both criminals and the criminal justice system are not completely deterministic. Mistakes, a lack of information, or criminal behavior that is not influenced by sanctions could all result in the threats being carried out, even if only infrequently. Even if complete deterrence of some acts is feasible, severe punishments may have additional costs to administer or monitor. Crimes that have severe punishments, and crimes that might be mistaken for a crime with a severe punishment (e.g. manslaughter may be difficult to distinguish from murder), should be more carefully investigated and more intensely litigated. As long as threats themselves have a potential cost, complete deterrence of offenders who are responsive to sanctions may be possible but not optimal.

We extend Becker's work by modeling a representative agent who chooses how much of each of n different activities to engage in. The individual is restricted by a temporal budget constraint with total time T. The activities are indexed $\{1, 2, ...n\}$. The individual is a utility maximizer with preferences represented by a strictly quasi-concave, differentiable utility function. For each activity, j, she faces a time cost of t_j and must decide what quantity of activity, x_j, to undertake. In addition to the time cost t_j, the government may impose sanctions for some of the activities. Following Becker, the punishments chosen by the government are described by two parameters for each activity j. s_j is the sanction per unit of activity j that the government detects, and p_j is the probability of detection for activity j. The individual is risk neutral and thus responds only to changes in the expected time cost, including punishment, of each activity.[7] The individual's problem can be written:

$$\max_{x_j \geq 0} U(x_1, ..., x_n) \text{ subject to } \sum_j (p_j s_j + t_j) x_j \leq T$$

The solution to the individual's problem can be characterized by a set of demand functions $\mathbf{x}^* = \{x_1^*, x_2^*, ...x_n^*\}'$. Each demand function depends on the total time the individual has and the time prices she faces: $x_j^*(\sigma_1, ..., \sigma_n, T)$ where $\sigma_k = p_k s_k + t_k$. We omit T from the demand function because we assume that it is constant.

[7] Because we model a representative agent and mean sanctions, the representative activity choices will not violate the budget constraint. Any given individual may violate her budget constraint depending on how often her criminal activity is detected. This is analogous to models of lifetime consumption decisions under uncertainty, in which a person could die with unspent wealth or unpaid debts.

The government, too, must solve an optimization problem. Its problem is to select a set of $2n$ parameters such that total social cost is minimized. The government faces three varieties of cost that combine to generate a cost function, C: (1) the direct cost of crime; (2) the cost associated with the vector of sanctions; and (3) the cost associated with detection. C is increasing in all of its arguments. The government's problem is:

$$\min_{p_j, s_j \geq 0} C(\mathbf{x}^*(\sigma_1, ..., \sigma_n), \mathbf{p}, \mathbf{s})$$

We assume that the cost function has second order conditions such that the local optimum is the unique global optimum. The government's optimal policy is fully characterized by the first order conditions. The first order conditions are:[8]

$$\begin{aligned} 0 =& \frac{\partial C}{\partial p_k} + \sum_j \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} \frac{\partial \sigma_k}{\partial p_k} \\ 0 =& \frac{\partial C}{\partial s_k} + \sum_j \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} \frac{\partial \sigma_k}{\partial s_k} \end{aligned}$$

The first order conditions carry the intuition that the 'cost' of increasing s_k or p_k must equal the 'benefit' which is the fall in costs due to the decrease in all crime. The conditions can be used to solve for the optimal s_k and p_k as follows, assuming an interior solution:

$$\begin{aligned} s_k &= -\frac{\frac{\partial C}{\partial p_k}}{\sum_j \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}} \\ p_k &= -\frac{\frac{\partial C}{\partial s_k}}{\sum_j \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}} \end{aligned}$$

We define the elasticity of a crime rate with respect to enforcement against that crime as an *own price elasticity* and the elasticity of a crime with respect to another crime's enforcement as a *cross price elasticity.*

2.2.2 Complements and Substitutes

We can distinguish activities that are complements, $\frac{\partial x_j^*}{\partial \sigma_k} < 0$, and substitutes, $\frac{\partial x_j^*}{\partial \sigma_k} > 0$, with activity k and separate the above summand. We use the notation $\sum_+$ to refer to the

[8] We write $\frac{\partial C}{\partial p_k}$ to denote the partial derivative of C with respect to the second argument; we write $\frac{\partial C}{\partial s_k}$ to denote the partial derivative of C with respect to the third argument.

sum over all activities j that are gross substitutes with activity k, that is $\sum_{+} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} = \sum_{j \text{ s.t. } \frac{\partial x_j^*}{\partial \sigma_k} > 0} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}$. The notation $\sum_{-}$ refers to the analogous sum over all of the activities that are gross complements with activity k.

$$\sum_j \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} = \sum_{+} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} + \sum_{-} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} + \frac{\partial C}{\partial x_k^*} \frac{\partial x_k^*}{\partial \sigma_k}$$

Activity k should be punished whenever:

$$\sum_{-} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} + \frac{\partial C}{\partial x_k^*} \frac{\partial x_k^*}{\partial \sigma_k} < -\sum_{+} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}$$

even if k has no social cost (i.e. $\frac{\partial C}{\partial x_k^*} \leq 0$). Similarly, the optimal punishments sometimes leave socially harmful activities unpunished. Activity k should go unpunished whenever:

$$\sum_{-} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} + \frac{\partial C}{\partial x_k^*} \frac{\partial x_k^*}{\partial \sigma_k} > -\sum_{+} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}$$

even if k is socially harmful (i.e. $\frac{\partial C}{\partial x_k^*} > 0$). In fact, when the model is modified to allow for subsidies, the optimal sanctions may include a subsidy for certain criminal activities.

2.2.3 Comparison to the Naive Social Planner

The difference between our model and prior literature is the harm weighted sum of the response of related crimes: $\sum_{+} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k} + \sum_{-} \frac{\partial C}{\partial x_j^*} \frac{\partial x_j^*}{\partial \sigma_k}$. We refer to this term as the correction for related crimes. When this term is positive (negative), activity k is a harm weighted gross substitute (complement) to other activities. Consider the behavior of a policymaker who does not recognize that crimes are interrelated. This 'naive policymaker' and the associated 'naive punishments' do not correctly maximize the social welfare function. The optimal punishments are more (less) harsh than the naive punishment when an activity has a positive (negative) correction for related crimes.

2.2.4 Relationship to Pigouvian Taxation

These punishments bear some resemblance to a Pigouvian tax. We model incarceration as a time tax designed to disincentivize socially harmful behaviors. The classic Pigouvian taxation result states that the tax is equal to the marginal social harm at the optimal activity level. In our model the first-best activity levels are generally not optimal because administering punishment is costly (Kaplow, 1990*a*; Polinsky and Shavell, 1982). Within this second best environment the optimal punishments are partly a function of spillovers from other externality producing activities.

By assumption every crime has a net social cost. Thus it is not optimal to set punishments sufficiently high to achieve first best levels. Instead the optimal activity levels are higher than their first best alternatives. Any optimal 'tax' in a setting with second best activity levels will be based on a weighted sum of other activities' cross price elasticities and the externalities of those activities.

The features of our model are prominent features of the criminal justice system but are also present in any environment in which there is regulation of externality generating activities. For example, administrative cost, measurement error, and political power will likely make it suboptimal to induce the first-best activity levels for polluting activities. If this is the case, regulations and taxes should be set taking into account the spillover effects to other externality causing activities.

2.3 Applications of the Model

As mentioned in the introduction, legislatures are aware that crimes are interrelated.

"Simply put, prescription drug abuse is one of the fastest growing drug problems in the nation, resulting in ever increasing rates of robberies and other attendant crimes." - US Senator Sherrod Brown advocating for enforcement against theft of prescription drugs

"Illegal aliens commit horrendous crimes against American citizens, crimes that strain State and Federal judicial systems, police and sheriff departments, and prisons that are already overcrowded and in a financial crisis." - US Representative Mo Brooks discussing immigration reform

However, there is no evidence that interrelation is taken into consideration carefully or consistently. Although it is probably true that some immigrants commit other socially costly crimes, it is also likely that some undocumented immigrants avoid crimes with small payoffs because those crimes carry the additional risk of being caught for undocumented immigration. While punishing undocumented immigration may incapacitate some individuals who are committing a variety of crimes, strict enforcement against undocumented immigration will decrease the marginal deterrence of increased participation in other criminal activities.[9]

[9] This is an application of marginal deterrence, originally discussed by Stigler (1970), to our model of multiple crimes. Whether enforcement of immigration law would increase undocumented immigrant's participation in criminal activities is ultimately an empirical question. For a discussion of the existing evidence,

More generally, both unpunished substitutes and harshly punished complements exist in the US criminal justice system. For example, drug possession by itself, as opposed to drug use, causes no social harm. Drug possession is punished only because it is a complement with other harmful activities. Similarly, the punishments for carrying a concealed weapon are best explained by the argument that carrying a weapon is related to using it in a harmful way. The controversy around concealed carry laws is at least in part a controversy about whether concealed carry increases or decreases other socially harmful activities. There are numerous studies on both sides of this issue. For example, Lott and Landes (2000) consider whether more relaxed carry laws might increase the chance that others have guns and thereby deter shootings. On the other hand Duggan (2001) links higher rates of gun ownership to additional gun homicides.

In many cases our model suggests a logic underlying existing punishment schemes. For example, if drugs tend to be substitutes, our model explains why it may be optimal to criminalize most drugs, while leaving a few drugs, such as alcohol and tobacco, uncriminalized. This policy is optimal if for relatively low enforcement cost it steers users towards a select few drugs. The more harmful a drug, the better a candidate it makes to be criminalized. The more inelastic its response to its own punishments, and the larger the cross price elasticities with costly activities, the better a candidate a drug makes to remain legal. Free needle exchanges take this reasoning one step further. If using a clean needle is a near perfect but less socially harmful substitute for sharing needles, then subsidizing a needle exchange is socially optimal.

Looking to medieval English law we find another interesting case. Pollock and Maitland (1899) state that burglary could be excused if committed by a hungry man whose aim was to steal a small amount of food. A sufficiently hungry man's criminal acts to acquire food might be unresponsive to sanctions. The distinction between 'theft for food when hungry' and 'other theft' carves out an opportunity for the hungry man to commit a crime for food without society being forced to incur the costs of sanctioning the theft while retaining strong disincentives for other acts which cause more net social harm.

The examples in this section suggest that legislatures, judiciaries, and prosecutors already take into consideration crime interrelation. Nonetheless, some might find it odd that a person be punished for a socially harmless activity or that the punishment of one activity be based on the commission of another. This, however, is not as striking as it may initially seem. Note first that the model does not suggest a violation of due process. The criminal must perform an action to be sanctioned and a consistent sanction is applied to each punishable activity. Prior work by legal scholars discusses the significance of prosecutorial discretion

see Bell and Machin (2013).

when there is information about criminal activities other than the charged offense. Richman and Stuntz (2005) explain that such 'pretextual prosecution' is typically allowed under the argument that tax evasion, for example, is a legitimate crime and criminals should not be exempted from prosecution for tax evasion because they engage in more harmful illicit activities. Richman and Stuntz cite Wayte v. United States, 470 U.S. 598 (1985) as the leading case addressing this argument.

Second, the model does not require an expansion of punishment setting powers. Legislatures are free to map an activity to any sanction they like, subject to the constraint that the punishment not be cruel or unusual. If the sanction is a prison sentence and the criminal is a mentally competent adult, the legislature is given nearly free reign.

Third, the model accepts the utilitarian premise that punishment should be socially optimal. It follows that the effect of punishing activity A on the commission of crime B should be included in the optimal sanction. Outside of a utilitarian structure a punishment may be considered a moral construct that depends only on the act it punishes, but a utilitarian framework cannot accept an a priori punishment–fair, right, or just punishments make sense only insofar as they affect individuals' utility. It is inconsistent with utilitarianism to accept that a punishment should be designed to deter the crime that it punishes, but no other crime. The optimal utilitarian punishment must take into consideration all of its effects.

2.4 Empirical Evidence

A large literature attempts to relate policy variables, such as the death penalty, police force size, clearance rate, even gun ownership regulations and the existence of a state lottery, to crime rates. Some of these studies measure responsiveness in a way that can be interpreted as an own price elasticity. Very few studies measure anything like a cross price elasticity. Some studies use control variables which could proxy for a price of a related activity, but they do not systematically investigate the relationship. For example Mikesell and Pirog-Good (1990) examine the effect on crime rates of the existence of a state lottery. Corman and Mocan (2000) use Assistance for Families with Dependent Children as a control variable. One reason to include AFDC as a control variable is because it can proxy for the labor market opportunities of the individual. In this sense it is related to the 'price' of work in the mainstream labor market.

Kuziemko and Levitt (2004) calculate the the effect on the aggregate crime rate of imprisoning drug, property, or violent offenders. The effect of switching one type of offender for another, at the margin, is not statistically distinct from zero. This is interpreted as suggestive evidence that some agents in the criminal justice system are making adjustments in order to minimize the total crime rate subject to the constraint of a fixed prison system size. Minimizing the aggregate crime rate is unlikely to coincide with opitimal policy. As described earlier in the paper, the optimal policy does not treat each type of crime as equivalently costly.

The following two studies calculate cross price elasticities for a group of crimes. Hakim, Spiegel and Weinblatt (1984) calculate cross price elasticities for different property crimes using a set of simultaneous equations. Levitt (1998*b*) calculates the elasticities of index crimes with respect to arrest rates for the crime itself as well as the elasticity with respect to an aggregate of other violent crime arrest rates and an aggregate of other property crime arrest rates.

Study	Crimes studied	'Prices'	Result
Levitt (1998*b*)	Index crimes	Arrest rates, arrest rates for groups of other index crimes	Cross price elasticities negative and about half the magnitude of own price elasticities; negative elasticities $\in [-0.012, -0.298]$, robbery $+0.062$ with respect to pecuniary crime arrest rate; larceny $+0.048$ with respect to violent crime arrest rate
Hakim, Spiegel and Weinblatt (1984)	Auto theft; burglary; larceny; robbery	Clearance rates	Negative own price elasticities $\in [-0.171, -0.891]$; most cross price elasticities are positive $\in [-0.329, 1.444]$

The cross price elasticities calculated by Levitt (1998*b*) are about half as large as own price elasticities. An important question to ask is how significant these cross price elasticities are. A simple interpretation would be that cross price effects should account for about half of punishments. Below we detail more precise approaches to determining the importance of cross price elasticities.

First, we estimate the potential gains from reallocating existing police resources across crimes. The computational method uses existing data to find the arrest rates which minimize the aggregate harm to victims of index crimes while keeping enforcement costs constant. Our approach is motivated in part by the availability of information about the cross price effects of crime.

2.4.1 Sources of Data

We consider only index crimes and use data about the number of each type of offense and the number of arrests for each type of offense as reported in the FBI Uniform Crime Reports (UCR). The elasticities of crime rates with respect to arrest rates come from Levitt (1998*b*), and the reported costs for each additional crime come from McCollister, French and Fang (2010). These costs include per crime cost to the criminal justice system, lost productivity of offenders, tangible costs to the victim and intangible costs to the victim. The estimates reported in McCollister are $8.98 million per murder, $241 thousand per rape, $42 thousand per robbery, and $107 thousand per assault. The harm for each burglary is $6462. The harm per incident of larceny is $3532 and the harm of auto theft is $10772 per offense.

2.4.2 Computational Method

Levitt (1998*b*) reports elasticities of index crimes with respect to index crime arrest rates. We convert the elasticities reported by Levitt to an elasticity with respect to arrests. Arrests for crime i, a_i, can be interpreted as the 'price' of crime to the individual[10], and the reported number of each crime, x_i can be interpreted as a 'demand' for crime. The demand for each crime has an elasticity with respect to each category of arrest. Assuming that these elasticities are constant over the relevant range of arrests, these are the demand functions describing the amount of each crime that occurs for any given combination of arrests:

$$x_j(\mathbf{a}) = K_j \prod_i a_i^{\varepsilon_{ji}}$$

where K_j is a constant determined by the observed values of x_i and a_i. ε_{ij} is the price elasticity of demand for crime j with respect to price i.

We assume that the costs per offense, B_j, reported in McCollister, French and Fang (2010) are constant in the relevant region, so the total harm to victims is $\sum_j B_j x_j(\mathbf{a})$. We compute the cost for an arrest of each type of crime, and assuming that this cost A_i is constant in the relevant region, the problem of finding the arrest rates that minimize the harm to victims while holding enforcement costs constant is:

$$\min \sum_j B_j x_j(\mathbf{a}) \text{ subject to } \sum_i A_i a_i = \bar{A}$$

To compute the optimal arrest rates, we test combinations of arrests which satisfy the constraint $\sum_i A_i a_i = \bar{A}$ and are within a certain range of the 2010 arrest rates. We then use

[10] For these computations we ignore the length of sanctions.

the demand functions and per crime costs to compute the harm to victims associated with each of those combinations of arrests. These arrest rates map to crime levels which then map to total victim harm. We record the combination of arrests that yield the lowest total victim harm and then repeat the computation for arrests that are similar to that combination of arrests. This algorithm terminates when no tested vector of arrests lowers total victim harm.

2.4.3 Estimates of the Harm Reduction

The 2010 crime levels correspond to a total victim harm of approximately \$296 billion. Assuming that the current arrest rates are optimized using only own price elasticities (i.e. assuming the naive optimum) and keeping enforcement cost constant, using cross price elasticities reduces the harm to \$288 billion. This represents a 3% reduction in harm to victims by reallocating existing enforcement resources.

A similar exercise minimizes total victim harm allowing enforcement expenditures to vary. Because the cross price elasticities reported in Levitt are mostly negative, this approach causes enforcement expenditure to rise. The reduction of harm to victims is about \$34 billion but is offset by a \$19 billion increase in enforcement costs resulting in a net reduction in total cost of crime of \$14 billion. These values are associated with large changes in levels of enforcement. Assumptions about the constant marginal harm of crime, the constant marginal cost of enforcement, and the constant Slutsky matrix are less tenable for large changes. Nonetheless, this suggests that changing enforcement levels based on cross-crime elasticities can have significant benefits in addition to the benefits of reallocating existing enforcement resources.

These estimates are conservative. These estimates only account for the seven index crimes and one aspect of expenditure on criminal justice. The harm from non-index crimes, wrongful convictions and other social burdens are not factored into these estimates. Additionally, Levitt groups crimes when making his estimates and his econometric specification suffers from division bias. We expect both of these factors to artificially reduce the variance in responsiveness across activities. The welfare impact of reallocating existing resources is driven by the variation in cross price elasticities, relative to own price elasticities. There may be a larger effect if the elasticities were computed for each crime arrest pair.

2.4.4 Estimates of Marginal Effects

An alternative calculation of the significance of the cross price elasticities considers the effects of a small change from current enforcement levels. For example, a small increase in the enforcement against assault will decrease assaults. It will also affect other crime rates.

The elasticities of each crime with respect to the arrest rate for each other crime allows us to calculate the change in the number of crimes that would occur in response to a change in each crime's arrest rate.

Using the UCR data and Levitt's elasticities we can calculate the change in the number of each index crime that would result from a 1% increase in the arrest rate for assault. Such an increase would lead to approximately 1306 fewer assaults according to Levitt (1998*b*).[11] Using Levitt's estimates the increase in assault arrest rates would decrease the number of murders by 14.3 and decrease the number of rapes by 6.6 but the change would increase the number of robberies by 213. The changes in property crimes would include 193 fewer burglaries, 1331 fewer larcenies and 301 fewer auto thefts.

Using per crime cost estimates from McCollister, French and Fang (2010), we find that the assault arrest rate change described above would reduce the harm to victims by about $270 million. The decrease in assaults alone would be valued at $140 million, while the decreases in the number of murders and rapes would correspond to a decreased burden of $129 million and $1.5 million. The increase in robberies would increase the harm to victims by $9.0 million. The decrease in burglaries larcenies and auto thefts would correspond to reduced harms valued at $1.2 million $4.7 million and $3.2 million.

The significant insight from this calculation is that approximately half of the harm reduction is due to a reduction in crimes other than assault. In this example, the decline in murder represents a large fraction of the harm reduction from a marginal increase in arrests for assault. The actual benefits of raising the arrest rate for assault is almost 2 times as large as would be calculated by a policymaker who did not include the effect on other crimes. When we use a small negative value for the own price elasticity of murder, we find that cross crime effects only account for 10% of the marginal harm reduction. However, if we use Levitt's measured, positive own price elasticity for murder, then increasing the arrest rate for murder actually increases the net harm to victims. If increasing arrests for murder increases the number of murders, then the cross crime effects are the only gain from increased arrests for murder. Repeating the calculations for the other arrest rates we find that cross crime effects account for between 29% and 67% of the harm reduction at the margin. Refer to the table in appendix VI.3 for more details.

[11] The cross price elasticities reported by Levitt are for groups of crimes. To use the example of murder and nonnegligent manslaughter, Levitt reports the percentage change in the number of murders associated with a 1% change in the 'arrest rate for violent index crimes excluding murder'. To calculate the change in the number of murders we first calculate the change in the 'arrest rate for violent index crimes excluding murder' which occurs when the arrest rate for assault rises 1%. In this example, the change in assault arrest rates causes the 'arrest rate for violent index crimes excluding murder' to rise by roughly 0.623%. Given the reported cross price elasticity of -0.129 and 14,722 murders in 2010, we calculate the decrease in murders reported above.

2.5 Conclusion

This paper generalizes prior work on criminal punishments by allowing multiple interrelated crimes. The insights extend beyond crime to any environment with spillovers between socially harmful or beneficial activities. When a friction makes first best activity levels suboptimal, the optimal policy intervention accounts for spillover effects of other externality generating activities. Crime is an excellent application since the most common sanction, incarceration, is sufficiently costly to ensure that the activities are not set to their first best levels.

The main intuitive contribution is that complementary (substitutable) crimes should be punished more (less) harshly than they would be if punishments were set for each crime in isolation. Moreover, making few assumptions, we provide a precise formulation for how to find the optimal punishments. The optimal punishment is proportional to the sum of the responsiveness of each crime with respect to that punishment, weighted by the measure of the costliness of that crime. A corollary of this insight is that social harm is neither sufficient nor necessary for an activity to be punished under the optimal criminal justice policy. The optimal punishments are based on the relationships with all activities, whether they are criminalized or not and whether they are harmful, beneficial, or benign. Some harmful activities should not be punished at all because of their substitutability, while other benign or beneficial activities should be punished because they are complements with harmful activities.

Applying this insight to criminal punishments promises significant benefits to society. For index crimes alone and using conservative assumptions, current enforcement spending could be reallocated to reduce the harm to victims by approximately $8 billion, or about 3% of the total harms to victims.

Other possible practivcal applications include the taxation of pollution, hiring based on big data (e.g. using zip code as a signal for future employee retention), taxation of unhealthy foods, subsidies for public health projects, and other principal agent problems with multiple activities.

CHAPTER III

The Welfare Impact of Corporate Tax Privacy

Abstract

Under Internal Revenue Code, Section 6103, most of the information contained in corporate tax returns is not publicly available. This paper investigates what corporations would do if they had access to other corporations' returns, what investors would do if they had access to corporate returns, and ultimately how these behavioral responses would affect welfare. The analysis suggests that corporate tax preparation and sheltering technology would become more widely available as firms learned from each other's returns. This would shift investment away from firms that have relatively good tax preparation and sheltering technology and toward firms that are relatively more productive. Socially wasteful expenditure aimed at lowering effective tax rates would also fall. Tax rates would likely need to rise in order to maintain government revenue, but the increase in productivity and decrease in socially wasteful expenditure would be welfare improving. The additional information that investors would gain would improve investors' estimates of the returns and risks of investing in each corporation, which would also be welfare-improving.

Keywords: Allocative efficiency, Behavioral response, Corporate income tax, Corporate tax return, Investment, Rent-seeking, Welfare
JEL Codes: H22, H25, H26, K34

In 2011, the New York Times published an article about General Electric's United States corporate income taxes, claiming that GE had paid no US tax bill despite having had billions of dollars of income (Kocieniewski, Mar. 24, 2011). The article's shockwaves reverberated throughout the public sphere. Congressman Paul Ryan requested information from GE that he presented at a town hall meeting, eliciting gasps, and numerous pundits weighed in (McCormack, Nov. 17, 2011). As it turned out, GE's corporate income tax was not quite as scandalous as first reported: GE's particularly low effective tax rate could be largely explained by the net operating loss carryforward that resulted from the $30 billion collapse of GE Capital during the Great Recession (McArdle, Mar. 29, 2011).

Nevertheless, as the various journalistic and political inquiries revealed, GE is a paragon of tax sheltering. GE's tax department employs nearly 1,000 people and is sometimes referred to as the world's best tax law firm. Moreover, the 2010 tax return that set off the hullabaloo was approximately 57,000 pages long—a testament to GE's intricate and innovative tax planning (Kocieniewski, Mar. 24, 2011).

While GE has received considerable attention for its tax prowess, some of the consequences of its sheltering behavior have not yet been fully explored. First, if GE is able to shelter its income better than other companies, then it should be a relatively more attractive investment—even compared to more productive companies if those companies are materially worse at tax sheltering. Second, if there are many firms that, like GE, are tax sheltering, they might all be independently and redundantly expending resources to develop the same tax shelters. Third, if GE's profitability is dependent on its tax decisions, then investors trying to determine how valuable GE is are severely handicapped if they are unable to see what GE's tax decisions are (Pearl, Sep. 21, 2016).

These are three economic distortions that arise from the intersection of tax complexity and tax privacy. The more difficult tax information is to acquire, the greater the distortion from misallocated investment, socially wasteful expenditure, and firm value uncertainty. If the corporate tax were as simple as possible (i.e. a lump sum tax), then (1) corporations could not meaningfully have differentiated tax preparation and tax sheltering technologies, (2) corporations would not spend socially valuable resources to develop new tax technology, and (3) investors would have no difficulty forecasting future tax liability.

Several policy prescriptions might alleviate these distortions. A simpler Tax Code could limit the potential for tax sheltering, decrease the incentive to invest in tax shelter development, and make forecasting future tax liability easier. A lower corporate tax rate would reduce the incentive to tax shelter and would decrease the extent to which tax decisions determine firm value.

This paper suggests an alternative to major tax reform to lessen these distortions: tweak the tax code to make the entire corporate return public. Drawing on tax law, financial accounting, and economic theory, this analysis attempts to determine what corporations and investors would learn from corporate returns and what their most likely behavioral response to that information would be. Ultimately, the objective is to determine how investment decisions would change in response to corporate tax information and how those changes in investment would affect welfare.

Whether corporations in the United States should be entitled to tax privacy is a question as old as the corporate tax. In 1909, Congress imposed a one percent excise tax on corporate net income in excess of $5,000. Under this law, corporate tax returns would live in the public record and be subject to public inspection. The overarching justification for the public record provision was that it would aid in the regulation of corporations—this was the primary appeal of the law according to President Taft. In 1910, the Commissioner of the IRS ruled that since Congress had not appropriated sufficient funds, the IRS was not obligated to maintain corporate tax returns in a searchable fashion. Eventually Congress did appropriate funds but simultaneously restricted corporate tax inspection to those authorized by order of the President (Kornhauser, 1990; Lenter, Slemrod and Shackelford, 2003; Blank, 2014).

In a provision of the Revenue Act of 1924, Congress made the amount of income tax paid by individuals and corporations public information. However, objectors—including President Coolidge—raised concerns about the privacy of individuals' information. In 1926 a revised revenue act made it such that only names and addresses, and not taxes paid, were public record (Kornhauser, 1990).

In the early 1930s, a Congressional investigation brought to light significant corporate evasion and income underreporting. Congress responded by including a provision in the Revenue Act of 1934 that required individuals and corporations to attach a pink slip to their returns. The pink slips were public information and recorded the taxpayer's name, address, gross income, amount of deductions, net income, and tax liability. Again, objectors raised privacy concerns and also argued that revealing proprietary information on tax returns could harm corporations. In protest, constituents sent fake pink slips to their Congressmen, and Congress repealed the law before it took effect (Pomp, 1993).

Since 1976, Section 6103 of the Internal Revenue Code has ensured that returns remain confidential, with some exceptions including (1) large shareholders, (2) certain government employees for tax administration purposes, and (3) government agencies pursuing a non-tax criminal investigation. Although statute protects corporate tax privacy, there is no constitutional right to tax privacy.[1]

When the corporate tax was initially contemplated, the opponents of corporate tax pri-

[1] *United States v. Dickey*, 268 U.S. 378 (1925)

vacy argued that public access would aid corporate regulation. Proponents countered that public access would constitute an invasion of privacy (Kornhauser, 1990). Since then many new lines of argument have emerged, some of which focus on the potential behavioral responses to publicly accessible corporate tax information (Lenter, Slemrod and Shackelford, 2003; Blank, 2014). This paper's analysis descends from that behavioral response lineage, but is distinguishable from its predecessors in two ways. First, it pays particular attention to the behavioral responses of investors. Previous work has focused on the behavioral responses of citizens, consumers, and corporate managers. Second, this analysis looks beyond what would happen to government revenue and tries to determine what would happen to the productivity of the economy.

If corporations had access to other corporations' tax returns, then they could learn from those returns. This learning process would make tax return preparation and tax sheltering technology more uniform across firms. Uniform tax preparation and sheltering technology would have two benefits. First, as the difference between firms' tax preparation and sheltering technology decreased, investors would choose to invest in productive firms instead of firms that had better tax preparation and sheltering technology. This would increase aggregate productivity. In other words, because investors make investments seeking after-tax profits, differentiated tax preparation and sheltering technology distorts capital allocation, shifting it away from relatively more productive firms to firms that have relatively better tax preparation and sheltering technology. Second, as firms gained easier access to tax preparation and sheltering technology, each firm would spend less developing those technologies—each would benefit from the advances of all the others. This would lead to less socially wasteful, redundant expenditure.

If investors had access to corporate returns they would be able to more precisely estimate future tax liability. More precise future tax liability forecasts would improve the precision of future net income and cash flow forecasts, which in turn would improve the precision of investors' estimates of firm values. Reducing uncertainty over firm values would allow investors to choose more profitable investments and find investments that matched their risk appetite, which would increase welfare. Furthermore, if tax policy is benevolent, then it is designed to induce a behavioral response from investors. For example, benevolent tax policy should place higher tax burdens on firms that generate negative externalities to discourage investment in those firms. The more difficult it is for investors to know what current tax liability is and what future tax liability will be, the more muffled the intended behavioral response will be.

The reason that corporations and investors can learn something from seeing corporate tax returns is something that this analysis refers to as tax variation. There are numerous sources of tax variation, including that different firms are eligible for different credits and that firms vary in how well they tax shelter. In some cases, tax variation is desirable—for example when a firm generates an externality that can be corrected with a tax. However, in other cases tax variation cannot be linked to a benevolent tax policy—for example if corporations have different tax preparation and sheltering technology. Moreover, almost every firm experiences tax variation from year to year, making it nontrivial for investors to forecast future tax liability.

This analysis is focused on corporate tax privacy and does not stray to address other important and related topics. In particular, whether there should be a corporate tax, whether the corporate tax should be territorial or worldwide, and what corporate tax rates should be are matters not explored here. This exercise simply posits that there is a corporate tax and that it will generate a material fraction of government revenue.

It is also worth noting that this analysis does not advocate tax planning and sheltering per se. Rather, it suggests that if tax planning and sheltering are already common, then it is better to lower the barriers that prevent learning about tax planning and sheltering. This prevents firms that are particularly good at tax planning and sheltering from having an economic advantage over other firms and it limits the wasteful, redundant spending on securing tax planning and sheltering advantages.

Lastly, this paper makes the case that eliminating corporate tax privacy would improve welfare, but it is unlikely that making corporate tax returns public would be revenue neutral. Corporations learning from one another could lead to additional government revenue (if, for example, firms learned to keep away from aggressive strategies that the IRS might combat); and investors learning from corporate tax returns could also lead to additional revenue (if, for example, investors flocked to firms with conservative tax approaches). The more likely outcome is that ending corporate tax privacy would require a policy change along another dimension to maintain government revenue. Changing other policy dimensions to raise additional government revenue will most likely create new economic distortions. In that case, the increased welfare of eliminating corporate tax privacy would have to be weighed against the new distortions to determine the total welfare consequences.

The remainder of the paper is arranged as follows. Part I examines how this analysis fits into the debate over corporate tax privacy. Part II provides a theoretical framework for the analysis. This framework posits some principles about investor behavior and also derives two important tax corollaries. Part III explores the sources of tax variation. Part IV investigates how corporations would use the information on other corporations' tax returns, what the effect on tax variation would be, and ultimately how welfare would change. Part V looks at how investors would use corporate tax information to improve their forecasts of the risks and returns associated with investing and how this would change welfare. Part V is followed by a conclusion.

3.1 The Corporate Tax Privacy Debate

The debate over whether corporate tax returns should be public is alive and well in the academy and occasionally makes an appearance in legislative circles too. The following table shows many of the arguments for and against corporate tax privacy.

Table 3.1 Corporate privacy debate

Arguments against privacy	Arguments for privacy
Education	Parity with individual
Shaming	Information overload
Increased detection	Proprietary information
Catalyzing policy changes	Reduced compliance
	Legal structure change
	Shelter proliferation and benchmarking
	Lost IRS strategic advantage

The following two subparts summarize the major arguments for and against corporate tax privacy. The third subpart explains why this analysis differs from previous research.

3.1.1 The arguments against corporate tax privacy

Education. Public access to corporate tax returns might increase interest in and understanding of tax law. However, even those who tout the educational benefits of tax publicity doubt that reducing tax privacy will lead to widespread understanding of the nuances of the tax system (Kornhauser, 2005). Moreover, many individuals do not increase their understanding of tax law, even in cases where it is in their best financial interest to do so (Bhargava and Manoli, 2015). It seems unlikely that public access to corporate tax returns will spawn a deep interest in tax law. Academics and others with interest in corporate returns, however, may learn a great deal from corporate returns.

Shaming. Public access to corporate returns might reduce aggressive corporate tax planning because investors, consumers, or some other group might remonstrate. While this is possible, there are reasons to doubt that the shaming effect will be large. First, not all avoidance strategies can be easily detected using corporate returns, especially by an untrained eye. Second, the majority of the evidence suggests that investors reward tax avoidance. For example, when Walgreens opted not to pursue an inversion when it merged with a foreign competitor, its stock price immediately fell over 14% (Hjelmgaard and McCoy, Aug. 6, 2014; Hanlon and Slemrod, 2009). Third, although consumers have shown an inclination to boycott corporations that engage in unseemly behaviors, this inclination is limited (Auger et al., 2003).

Increased detection. Public access to corporate tax returns would get more eyes on corporate tax returns and thus possibly increase detection of evasion. There is some evidence that tax compliance is higher for publicly traded firms and firms in heavily regulated industries, suggesting that non-IRS eyes can affect compliance (Rice, 1991). The IRS also encourages whistleblowers and informants by offering a reward of up to 30% of the collected proceeds. Furthermore, budget constraints are forcing the IRS to scale down some of its corporate audit programs, which might increase the likelihood that the IRS would benefit from outside help. However, while some experts might try to scrutinize returns, perhaps using statistical methods to find irregularities, the vast majority of people will not be of much help to the IRS—the requisite understanding of corporate tax will likely prove too high a barrier for most.

Catalyzing policy changes. Making more tax information public could increase pressure on Congress to make changes to the Tax Code. In particular, it might expose corporations that are benefitting from specific tax breaks, or it might increase the pace with which Congress makes changes widely considered to be beneficial. The Tax Reform Act of 1969, for example, was partially a response to a public outcry over the number of wealthy individuals that were using shelters to pay little or no tax (Wall Street Journal, Jan. 20, 1969; Shaviro, 1990).

3.1.2 The arguments for corporate tax privacy

Parity with individual. There are some who believe that as a first principle corporations and individuals should be entitled to the same rights under the law (Tax Executive Institute, 2006). However, there are meaningful distinctions between corporations and individuals that make the case for parity tenuous if not specious. First, there is a plausible public policy concern that individual tax return information could aid criminals interested in perpetrating various pecuniary crimes —not to mention the far more likely outcome that charity organizations would use the information to solicit contributions.[2] Second, the information on an individual return would reveal far more intimate details about the taxpayer (including

[2] 79 Congressional Record 2690 (1935) stating concern that conmen would use the individual return information to compile “sucker lists”.

potentially medical expenses, marital status, and number of dependents) than corporate tax returns would reveal about the various stakeholders in the corporations. Third, the relative availability of corporate financial information implies that social norms are less concerned with making corporate financial information available.

Information overload. Some believe that the information contained in corporate tax returns is so prodigious and confusing that releasing it would lead to either poor policymaking or poor corporate decision-making (Tax Executive Institute, 2006). This argument turns conventional wisdom on its head—normally, more and better information leads to better decision-making—and is unpersuasive. First, despite most people not knowing much financial accounting, providing access to financial statements is generally thought of as beneficial. It is not obvious why tax return data should be different. Second, the key players in most policy and corporate decision-making settings have expert advisors that are likely to mitigate any urge to succumb to confusion. Third, even if there were negative consequences to giving some people more information, it is highly unlikely that these negative consequences would outweigh the beneficial effects of academics, investors, and policymakers having access to more information.

Proprietary information. Some have made the claim that corporate tax returns contain information that, if released, could harm corporations (Tax Executive Institute, 2006). Although there are many pieces of information that a corporation might not want public, this argument specifically refers to information that other competitors could use to that corporation's disadvantage. While this raises a plausible issue, it seems unlikely that publicizing proprietary information on tax returns will cause severe economic disruptions—it may even be beneficial. Access to proprietary information is a two-way street. Firm A will benefit at Firm B's expense because A can see B's tax return, but B will also benefit at A's expense because B can see A's return. The proprietary information argument ignores this symmetry: competitors should both benefit from and be hurt by additional access to information. Access to proprietary information would also increase competitiveness, which in most cases is thought to be beneficial.

Reduced compliance. Corporations might reduce their compliance with tax law if the information they revealed to the government would then be revealed to the public (Tax Executive Institute, Jul. 25, 2002; Lenter, Slemrod and Shackelford, 2003). This is certainly a possibility—though it remains an open question how large this effect will be. If returns are made public, investors might reward firms for revealing more information, in which case increased compliance might be possible. If Congress and the IRS suspect reduced compliance, they could alter audit procedures and non-compliance penalties to minimize the negative effects of reduced compliance.

Legal structure change. Making corporate returns public might push some businesses to adopt non-corporate legal structures to avoid public disclosure (Lenter, Slemrod and Shackelford, 2003). In essence these companies would be reducing compliance legally by changing their legal organization. To the extent that this transformation would be costly or would cause companies to forgo benefits available only to corporations, this would create economic distortions. To mitigate this cost, new laws could create disclosure requirements for non-corporate businesses which themselves would likely be beneficial.

Shelter proliferation and benchmarking. With access to the universe of corporate returns, corporations would learn more about different tax preparation and tax sheltering technologies. Moreover, tax preparers (who would be partially freed from confidentiality agreements), activist investors, and consulting firms could catalyze tax shelter proliferation (Blank, 2014). This could lead to corporate tax standards for successful tax avoidance (i.e. benchmarking) (Lenter, Slemrod and Shackelford, 2003), and firms could be evaluated along avoidance metrics just as they could for customer satisfaction and workplace safety. Shelter proliferation and benchmarking would lower government revenue collected (Blank, 2014).

Lost IRS strategic advantage. With access to the universe of corporate returns, firms could assess which tax strategies are met with IRS resistance. With sufficient information, firms could estimate a tax aggressiveness frontier—essentially determining what combinations of tax strategies are safe from IRS action and which are not. If every firm moved to the maximum tax aggressiveness that would not invite additional IRS scrutiny, most firms would probably have lower effective tax rates. This would lower government revenue collected (Blank, 2014).

3.1.3 How this analysis fits in

The most persuasive arguments in favor of corporate tax publicity listed above are that experts could use the returns to study corporate behavior, potentially detect illegal behavior, and help direct the government towards better policy. The strongest argument in favor of corporate tax privacy is the potential for reduced compliance. At first, it may seem that the loss of government revenue due to tax shelter proliferation and lost IRS strategic advantage would be a drawback of corporate privacy. This analysis argues otherwise and suggests that, taking all things into consideration, the balance favors eliminating corporate tax privacy.

Corporate behavioral responses to tax return information will likely lead to lower government revenue—absent some change along another tax policy dimension. However, government revenue is not an end in itself. Levying taxes is only socially beneficial if the proceeds are used to improve people's lives or the tax discourages socially harmful activity. Because taxes distort incentives, the benefits of the tax must exceed the lost economic benefit caused by the tax. Moreover, since Congress may change many policy dimensions in the tax space, it should choose the tax system that is least distortionary.

The remainder of the paper argues that protecting government revenue by maintaining corporate tax privacy (1) distorts capital allocation, (2) encourages wasteful spending on developing tax-sheltering technology, and (3) complicates investor decision-making. Thus corporate tax privacy is unlikely to be socially beneficial and is almost certainly not the best method to maintain or raise government revenue.

3.2 Investor Behavior

The ultimate objective of this analysis is to determine the welfare implications of making corporate returns public by looking at how eliminating corporate tax privacy would change investment, limit socially wasteful spending on designing new tax shelters, and influence investor decision-making. This section discusses the aspects of investor behavior that are relevant to the welfare analysis and establishes a theoretical framework for use in later sections.

3.2.1 Basic principles

The literature on investor behavior is extensive, and at its heart lies portfolio theory. Portfolio theory explores how an investor should allocate her capital across different investments. The purpose at hand requires only a few simple, intuitive principles.

Investors maximize risk-adjusted returns. The first necessary step to solving the investor's portfolio problem is defining an objective. The fundamental objective of any investment is a return. Ceteris paribus, an investor will prefer a higher return to a lower return. Returns are, however, not the only consideration—investments also vary in their risk. Making some assumptions, an investor can adjust the return of each investment by some factor that accounts for the risk. After this adjustment, the investor should aim to allocate her capital to maximize these risk-adjusted returns.

This procedure explains the majority of investing behavior from a hedge fund buying distressed debt at a discount hoping to collect through the courts, to a nonprofessional investor picking a few stocks as a hobby. Investors will have different beliefs over what will happen in the market, will vary in the sophistication of their methods, and will span the wealth distribution—but each is attempting to maximize her return, adjusting for risk.

Investors forecast future returns. Investors make investment decisions anticipating a stream of income. A firm may have a large or a small income next year, regardless of what its income is today. Thus, before an investor can solve her portfolio problem, she must know (or, more likely, estimate) the return and risk associated with each investment. The standard approach to this estimation posits that the world tomorrow could take any one of several states, each of which is assigned a probability. Consider the following table as an example:

Table 3.2GM return

	Boom	Bust
Probability	.5	.5
Return	$30	$10

The expected return is $20, and a frequently used measurement of the risk is the standard deviation, which is $10. A more complicated scenario might include several different stocks, many time periods, and multiple states of the world. In that case, the investor would have to forecast the possible returns for several periods for each stock and each state of the world.

Sophisticated investors perform this procedure using analytical methods and computers. Other investors may simply have a gut feeling about a particular stock. Both of these are forecasts and special cases of the procedure described above.

Information improves investor forecasts. To forecast future returns investors need information. The information could range from a firm's past financial statements, to management interviews, to an overheard conversation on a subway, to a broadcast of Jim Cramer's Mad Money television show. At one extreme, an investor might have insider information; at the other extreme, an investor might know little more than a stock's ticker symbol. Just as importantly, investors vary in their ability to use information. Some have intricate computer models, some are trained Form 10-K readers, and some have neither specialized equipment nor training. Without information, investors cannot make forecasts—even informal ones—in which case their investment decisions are random.

More information is not necessarily better. Misinformation is unlikely to improve an investor's decisions and even good information could result in information overload. However, for the purposes of this analysis, assume that additional information will improve an investor's forecast.

To explore how more information could improve an investor's forecast of future returns, consider the following example.

Table 3.3 GM limited information

Year	2015	2016
Income	$100	$140

An investor is attempting to forecast GM's 2017 income using the information in the above table. If that were all the information the investor had, she could make the reasonable assumption that the firm's income will continue to grow at the same rate of 40% and project $196 for 2017 income.

Now assume that the investor has the following information in addition to the income figures.

Table 3.4 GM additional information version 1

Year	2015	2016
Sales	$500	$550
Manufacturing Costs	-100	-110
Rent	-300	-300
Income	$100	$140

Note that in both years manufacturing costs as a percentage of sales were 20% and rent was $300. Now if the investor assumes that sales grow at the same rate, the cost of manufacturing as a percentage of sales remains constant, and rent remains $300, then the investor will project 2017 income of $184.

The first estimate, $196, would imply at least one of the following: (1) sales growth increased, (2) the cost of manufacturing cars decreased, or (3) the rent decreased. Since there is no basis for any of those assumptions, the second forecast is probably better.

In the previous example, the investor made a forecasting error when she was unable to distinguish between fixed and variable costs. Similarly, an investor would make a forecasting error if she were unable to distinguish between one-time and recurring items. Consider the following example.

Table 3.5GM additional information version 2

Year	2015	2016
Sales	\$500	\$450
Manufacturing Costs	-100	-90
Rent	-300	-300
Recurring income	\$100	\$60
Lawsuit income	+0	+80
Total income	\$100	\$140

Note that in all three examples total income is the same and that manufacturing costs are 20% of sales. Now if the investor assumes that sales fall at the same rate, the cost of manufacturing as a percentage of sales remains constant, and rent remains \$300, then the investor will project 2017 income of \$24. This is substantially different from the original estimate of \$196.

In both of these examples more information improved the forecast. In the first, the improvement came from forecasting more elements forward—essentially being able to distinguish between fixed costs and variable costs. In the second, the improvement came from being able to distinguish between one-time and recurring items. These examples used fairly simple forecasting methods, but the same applies to more complicated models. More relevant, accurate information used correctly will yield better forecasts.

3.2.2 Tax corollaries

The previous subpart established some general principles about investor decision-making. This subpart considers those principles through the lens of corporate taxes and derives two corollaries. In later sections these corollaries will establish a framework that will link tax return information to investment decisions.

Tax rates affect investment allocation. Since investors chase returns and taxes modify returns, taxes will in many cases affect investment. In particular taxes that are not uniform across firms will affect investment. By adding some structure to the relationship between investment and returns, the following model sheds light on the effect that differentiated tax rates have on investment.

An investor can invest in two corporations. She may divide her investment between the two corporations however she likes. Assume that each corporation's pre-tax income is an increasing function of the investment it receives. In other words, the more investment the firm receives, the more pre-tax income it generates. Assume further that marginal pre-tax income is decreasing for both firms. In other words, the first dollar of investment will generate more pre-tax income than the second dollar, which will generate more pre-tax income than the third dollar, so on and so forth. Consider the following table as an example.

Table 3.6GM investment schedule

Investment	Cars manufactured	Pre-tax income
$0	0	$0
1	10	10
2	18	18
3	24	24
4	28	28
5	30	30

Notice that pre-tax income increases with investment but that each additional dollar of investment generates less additional pre-tax income—the first dollar of investment generates $10 of pre-tax income, but the second dollar only generates $8. Assume a second car manufacturer has the following pre-tax income schedule.

Table 3.7 Ford investment schedule

Investment	Cars manufactured	Pre-tax income
$0	0	$0
1	9	9
2	16	16
3	21	21
4	24	24
5	25	25

The investor receives the entire after-tax income of the firm as an increase in share price and must decide how to allocate her investment. If there is no corporate tax, the investor will set her investment allocation so that the marginal return of both investments is equal. This means that an additional dollar of investment in either firm yields the same increase in pre-tax income. This allocation is optimal because at any other allocation the investor could shift a dollar from the firm with lower marginal return to the firm with higher marginal return and increase her income. To continue the example, if the investor had \$5 to invest, then she would have the following options.

Table 3.8 Investment possibilities

Investment in GM	Investment in Ford	Pre-tax Income
\$0	\$5	\$0+\$25=\$25
1	4	\$10+\$24=\$34
2	3	\$18+\$21=\$39
3	2	\$24+\$16=\$40
4	1	\$28+\$9=\$37
5	0	\$30+\$0=\$30

The investor's most profitable investment allocation is \$3 in GM and \$2 in Ford. Note that this also is the allocation that leads to the largest number of cars manufactured. Because it maximizes the number of cars produced given the fixed investment of \$5, this is the best investment allocation in terms of welfare.

If both firms faced the same tax rate, then there would be no change in the investment allocation. The tax would reduce the incomes derived from both firms by the same proportion. The investor's return would fall, but the number of cars produced would remain the same.

If firms had different corporate tax rates, the investor would maximize her return by allocating her investment such that the investments had equal after-tax marginal returns. If at a certain allocation both firms have the same pre-tax marginal return, the investor could shift investment from the high tax firm to the low tax firm and increase her income. Doing so would lower the pre-tax income of the investor, but decrease the corporate tax burden by an even larger amount. The larger the difference in tax rates, the more investment the low tax firm would receive and the less investment the high tax firm would receive relative to the no-tax scenario.

Returning to the example, assume that Ford has a tax rate of 10% and GM has a tax rate of 30%.

Table 3.9 Investment possibilities

Investment in GM	Investment in Ford	After-tax Income
$0	$5	$0.7+$25.9=$22.5
1	4	$10.7+$24.9=$28.6
2	3	$18.7+$21.9=$31.5
3	2	$24.7+$16.9=$31.2
4	1	$28.7+$9.9=$27.7
5	0	$30.7+$0.9=$21

Now the optimal investment allocation is $2 in GM and $3 in Ford. Note that this reduces the number of cars manufactured from 40 to 39. The tax variation led to a worse investment allocation. The total number of cars decreased by 2.5%, even though the total investment, $5, did not change.

This model can be extended in a number of ways. The above logic applies equally well when there are an arbitrary number of firms. In the taxless scenario, the investor should allocate her investment so that the marginal return of every corporation is the same. When the model is augmented with a corporate income tax, the investor should allocate her investment so that the marginal after-tax income of every firm is the same.

Even the first (and therefore the most lucrative) dollar invested in some firms will not yield a return sufficiently large to merit the investment. This return could be positive (but smaller than alternative investments), zero, or even negative. In that case, some firms will not receive investment.[3]

Tax information improves income forecasts. As a corollary to the axiom that additional information improves investor forecasts, additional tax information should also improve forecasts. Previous examples of information and forecasts demonstrated that additional information could improve forecasts of sales, costs, and ultimately income. The same principles apply here, except that now the objective is to forecast after-tax income. Tax

[3] A complete model would also incorporate how much an investor would be willing to invest. Because each investment has decreasing marginal return—in both the pre-tax and after-tax cases—the marginal return of the investor's total investment will also have decreasing marginal return. This aggregate marginal return curve characterizes the firms' downward sloping demand for investment curve. The model could also include a supply curve—a function that relates after-tax marginal return to the amount of investment the investor is willing to supply, an upward sloping curve. The reason the investment supply curve slopes up is that the investor must forgo either leisure or current consumption to increase investment—assuming that the marginal utility of leisure and current consumption are decreasing, the investor will need higher and higher returns to be induced to forgo more and more leisure and current consumption. The investment level and after-tax rate of return at which the two curves intersect would be the equilibrium.

information could improve projections either by improving the forecast of future tax liability or by improving forecasts of future pre-tax income.

To optimally allocate her investment, the investor must, to the extent possible, predict the future effective tax rate of each corporation. Not having the insight of managers, nor the foresight to anticipate Congressional action or future economic conditions, the investor will have a distribution of possible effective tax rates.

In some cases, corporations with low effective tax rates today will have low effective tax rates in the future. For example, firms in industries that have easy access to tax credits and deductions will have lower effective tax rates today and tomorrow. This also applies to firms that can consistently use effective tax sheltering technologies. In contrast, firms that, for example, have unusually large credits today will have different effective tax rates tomorrow. This is especially true for firms that find ways of front-loading their deduction claims, which causes not only an unusually low effective rate today, but also an unusually high rate tomorrow.

The following examples demonstrate some of the ways that tax information could improve investor forecasts. Assume that an investor can only observe Apple's income for two years and would like to project income for 2017.

Table 3.10 Apple limited information

Year	2015	2016
Income	\$80	\$120

If this is all the information that the investor is given, she might make the reasonable assumption that after-tax income will continue to grow at the same rate of 50%. In that case, her forecast for 2017 income would be \$180.

Now assume that the investor has the following information:

Table 3.11 Apple additional information

Year	2015	2016
Income before taxes	\$100	\$160
25% Corporate income tax	-25	-40
Retirement plan startup costs credit	+5	+0
Income	\$80	\$120

Assuming that income before taxes continued to grow at a 60% percent rate, 2017 income would be $192. The additional information changes the forecast, probably for the better, because it allows the investor to separate the recurring and non-recurring portions of tax liability.

In the Apple example, the investor is learning some information about the company from the tax information—in particular that Apple is creating a retirement plan. In a more sophisticated model, an investor might use this information to improve forecasts taking into consideration Apple's ability to recruit good employees because of the retirement plan and the increased costs associated with operating the retirement plan.

Here is another example of how tax information can improve forecasts of company sales. Assume that the investor only sees the following information about Disney.

Table 3.12 Disney limited information

Year	2015	2016
Income	$185	$184

She may assume that income will decline at the same rate and forecast 2017 income of approximately $183. Now assume that the investor sees the additional information in the table below—for simplicity, the only costs that Disney incurs are the state corporate taxes.

Table 3.13 Disney additional information

Year	2015	2016
Sales	$200	$200
10% California tax	-10	-12
5% Florida tax	-5	-4
Income	$185	$184

If the investor knows that sales subject to each state's tax jurisdiction are apportioned by the fraction of sales that occur in that jurisdiction, then she will be able to deduce sales in California and sales in Florida as follows.

Table 3.14 Disney additional information

Year	2015	2016
California sales	$100	$120
Florida sales	100	80
Total sales	$200	$200
10% California tax	-10	-12
5% Florida tax	-5	-4
Income	$185	$184

Assuming that California sales continue to grow at the same rate of 20%, they will be $144 in 2017. Assuming that Florida sales continue to decline at the same rate of 20%, they will be $64 in 2017. California taxes will be $14.4 and Florida taxes will be $3.2. Thus the projected total income for 2017 will be $190.4 —again this forecast is materially different than the forecast that uses only the income information.

The examples thus far have focused on the investor's forecast of expected return, but tax information can also inform about the risks associated with investing in a corporation. As a final example, assume that an investor only observes the following information.

Table 3.15GE limited information

Year	2015	2016
Income before taxes	$100	$100
Corporate income tax liability	-25	-20
Income	$75	$80

Knowing why the corporate tax is different in 2015 and 2016 is relevant for forecasting both future expected return and future risk. For example, if 2016 tax liability were smaller because of a one-time credit, then 25% is probably a better approximation of GE's future tax rate. If the lower tax were caused by a risky tax avoidance strategy, which the IRS would likely contest, then the best approximation of GE's future tax rate would depend on the probability that GE prevails over the IRS and the penalties GE would face if it did not prevail. With a risky tax avoidance strategy, the best forecast of GE's tax rate could be lower than 25%, but investing in GE would also carry more risk. If an investor cannot see why a firm has a tax rate that is different from the statutory rate, she will not only have a worse estimate of future tax liability, but also a worse estimate of future tax risk.

3.3 Tax Variation

The previous section posited a few principles about investor behavior and suggested some of the effects of the corporate tax system on investing. Specifically (1) investment allocation depends on the extent to which firms face tax variation and (2) investors can improve their forecasts using tax information. Both of these effects stem from an underlying assumption of tax variation. If there were no tax variation across firms, then taxes would have no effect on the relative profitability of investing in different firms. If there were no tax variation over time, then tax information would have limited forecasting value.

Indeed, there is significant tax variation across firms and over time. This part explores that variation and sets the stage for later parts which will examine how reducing corporate tax privacy will affect welfare by changing the consequences of tax variation across firms and over time.

3.3.1 What is tax variation?

Tax variation consists of everything that would cause two firms' average or marginal tax rates to differ. Potential causes of tax variation include differences in pre-tax income, credits, deductions, location of foreign subsidiaries, and tax preparation and sheltering technology. The simplest way to think about tax variation is as variation in effective tax rates. Anything that causes two firms to have different effective tax rates is a source of tax variation. Some firms are more tax savvy or operate within industries that have access to more deductions and credits—these firms will have lower effective tax rates.

This analysis divides the sources of tax variation into two groups, legal and corporate, which may be further partitioned as the table below shows.

Table 3.16 Sources of tax variation

Legal	Corporate
Benevolent policy	Error
Lobbying	Avoidance
Inter-governmental competition	Evasion
	Judgment
	Chance

The following two subparts describe each of these sources of tax variation in more detail. As an aside, the underlying cause of most tax variation is the Tax Code. It sets brackets for different levels of income and allows different credits and deductions for different types of economic activity. The Code's complexity also creates more opportunities for tax sheltering and creates space for corporate error, judgment, and chance to enter into the picture.

3.3.2 Legal variation

Law is an important source of effective tax rate variation. The corporate tax rate is not constant over income, and the observed variation in effective tax rates is far richer than the statutory rate, which averages close to 35% for most of the corporate income distribution. The welfare consequences of tax variation depend on the purpose of the statute that generates that tax variation. However, the raison d'être of a statute is not always clear—what one hails as benevolence, another may denounce as political favoritism. The aim of this paper is not to determine whether any specific statute is the result of optimal policy or lobbying, but rather to explore the consequences, assuming the reason for the statute is known.

Benevolent policy. There are several reasons why a policymaker acting in the interest of a country might want to enact tax variation. Some of these are captured in the study of what economists call optimal tax. Optimal tax theory is a normative branch of economics that aims to determine what tax rates or systems should be implemented in order to achieve a certain goal—normally welfare maximization. Examples of optimal tax policies include correcting externalities, minimizing the excess burden of the tax system, and stimulating business during recessions through tax cuts.

An externality arises when the private net benefit of an activity is not equal to the social net benefit of that activity. For example, buying a cellphone not only benefits the buyer, but also increases the size of the cellphone network making owning a phone more valuable for other consumers. The more friends with cellphones one has, the more valuable owning a cellphone becomes. Because the phone buyer—at least in the traditional model—only considers the benefit to himself, cellphones will be socially under-demanded. In other words, even after all the privately net-beneficial transactions have taken place, there are still additional socially net-beneficial transactions that could take place.

Arthur Pigou proposed the classic solution to the externality problem. He suggested taxing goods with negative externalities and subsidizing goods with positive externalities. Enacting this policy would increase the quantity demanded for positive externality goods and lower the quantity demanded for negative externality goods.

In the traditional model, a good-specific sales tax is used to correct externalities. However, the corporate tax could be used and arguably is used in some cases. The Tax Code, for example, allows corporations to both deduct and claim a credit for R&D expenditures. R&D expenditures generate research that may have social value beyond the new and improved products that the corporation doing the R&D is able to produce. Other entities will, perhaps with a time lag, also have access to the new technologies.

In addition to correcting externalities, a benevolent policymaker will raise revenue while incurring the smallest deadweight loss. Papua New Guinea, for example, has higher taxes on income derived from some natural resources. Because natural resources cannot be relocated to another tax jurisdiction, taxing natural resources induces relatively few behavioral responses. In contrast, a customer service center is highly mobile and would be likely to relocate to another jurisdiction if it faced high tax rates.

Generating revenue efficiently is not the only mandate of the legislature. The legislature can use tax policy as macroeconomic stimulus to vitalize a torpid economy. Depending on how this is done, it could cause effective rate variation—for example, by accelerating tax depreciation or offering certain credits not uniformly applicable to all firms. One recent example of this is the Economic Stimulus Act of 2008, which allowed firms to deduct bonus depreciation.

Tax policy also reflects societal preferences. For example, smaller firms (measured by income) have lower statutory tax rates, firms with domestic manufacturing are eligible for tax deductions, and firms that make investments in low-income housing or hire Native Americans are eligible for tax credits.

Lobbying. Differentiated tax and subsidy treatment is sometimes the outcome of political favoritism. The poster child of this phenomenon is farm subsidies, but several other industries also receive preferential treatment, including biodiesel, oil and gas, and mining.

Inter-governmental competition. Governments competing for economic activity offer tax breaks to specific firms or industries, which leads to effective tax rate heterogeneity. Theory and experience suggest mobile firms are particularly attractive targets. In 2004, for example, the United States modified the tax treatment of international shipping income, thereby incentivizing international shipping companies to remain American. States also engage in tax competition. In 2000, six states offered credits for film production, ten years later all but six offered such credits (Leiser, 2014). Michigan offered a film tax credit to movie production companies who performed substantial work in Michigan (Ernst & Young, Feb. 2011). The credit program attracted investment, but was terminated in 2015, at which time the film business in Michigan dissipated (Egan, Jul. 20, 2015).

3.3.3 Corporate variation

Legal provisions are not the only reason that effective tax rates vary across corporations. There are several sources of corporate variation, including error, avoidance, evasion, judgment, and chance. These are not always clearly defined and distinct, but the following demarcations offer a starting point: error and chance are unintentional, but error requires that available information was neglected or misused; avoidance and evasion are intentional; avoidance is at least plausibly legal; evasion requires concealing information from the tax authority and is illegal; judgment implies that reasonable minds could differ on the correct application of law.

Error, avoidance, evasion, judgment, and chance explain how even identical businesses, subject to the same laws, could prepare different returns. This is possible because corporations are intricate entities, and the Tax Code is prodigious. Corporations must record and categorize the multitudinous tax-relevant transactions they execute every day. That process is time consuming and requires discretion. The IRS estimates that taxpayers spend nearly six billion hours per year complying with tax law, 359 million of which hours are spent by corporations on Form 1120 (Taxpayer Advocate Service, 2008). Moreover, the size of the Tax Code combined with the Code's complex nature renders compliance hardly a simple matter of course. Different tax experts given the same information could easily prepare different returns (Slemrod, Mar. 3, 2005). When the incentive to minimize tax liability is added to the mix, actualized by tax practitioners of varying moral fiber and skill, we should be surprised whenever two businesses do manage to have the same effective tax rate.

Error. One reason that two otherwise identical corporations could have different tax returns is the prevalence of errors in corporate tax accounting (Bloomberg, Jan. 2015). Common information management mistakes include incorrectly manually entering tax relevant data, accidentally deleting Microsoft Excel formulas designed to compute taxes, and overwriting tax relevant data. Common rules-based mistakes include companies prematurely closing books, adjusting asset values for past-years without making all the necessary adjustments for following years, writing off business units that still have value, and prematurely expensing deferred compensation. The errors seem to stem from two sources: uninterested management and the limited input solicited from tax professionals in the design of the accounting systems.

According to Bloomberg (Jan. 2015), as a result of corporate tax error, 16% of firms suffered unfavorable adjustments, 11% missed tax breaks because of data issues, and 7% grossly miscalculated tax provisions. In some cases, these errors are detected and corrected and ultimately do not affect the economic position of the corporation. If, however, detected errors result in penalties or forgone benefits, then they will have an effect on the economic circumstances of the firm. In the case where these errors are never detected, they will also have an effect on the firm by either permanently increasing or decreasing shareholder value.

Avoidance, undisputed. Common legal strategies to reduce tax liability include carefully controlling the timing of transactions or preferring certain forms of financing. Nearing the end of a tax-year it may be advantageous to shift revenue recognition from the coming year to the current one if rates are expected to be higher in the following year. Interest from debt obligations is deductible, but dividend payments are not. From a corporate tax standpoint, this means that debt is preferred as a financing method.

Avoidance, disputed. There is a grey area, where the IRS and a corporation (or even different corporations) might disagree about whether a particular application of tax law is kosher or not. Common examples include transfer pricing, especially when intellectual property is involved, and intercompany short-term loans. Microsoft, for example, sold intellectual property to a subsidiary in a lower tax jurisdiction. The subsidiary then used the intellectual property to generate profit. The original sale is subject to US taxes and the subsequent profit is subject to foreign tax law—until the subsidiary pays a dividend to Microsoft. The price of the original sale effectively determines the allocation of taxable income between these two jurisdictions. Although there is uncertainty over the value of the intellectual property, Microsoft has a strong incentive to estimate the value to be as small as possible. Although the profits belong to the subsidiary, they may be loaned to the parent company for periods less than 60 days. Rolling over these loans HP, for example, was able to effectively repatriate billions (The Economist, Sep. 21, 2012).

Evasion. The IRS estimates that in 2006 there was $67 billion of underreporting corporate tax liability, some of which was evasion (Internal Revenue Service, Dec. 2011). Particular examples of evasion in 2015 include underreporting income and over reporting expenses (Internal Revenue Service, Oct. 2, 2015). Public and private corporations have different incentives to evade. Because evasion typically makes a business appear less profitable, public corporations may prefer to pay more in taxes to avoid disappointing shareholders with low publicized earnings (Blumenstein, Berman and Perez, May 2, 2003). Private companies may only care about the after tax cash flow and may therefore be more inclined to evade.

Judgment. Determining tax information requires judgment. For example, firms can expense bad debts, but deciding what proportion of debts will be uncollectable is not an exact science and reasonable minds may differ. On the financial statements, judgment is also at play in determining the size of the tax cushion—the provision that firms make in anticipation of dispute with the IRS in which they might not prevail.

Chance. Chance can also play a role in a firm's effective tax rate. Consider, for example, two identical firms. One makes a large capital expenditure on December 31, 2016, and the other makes the same large capital expenditure on January 1, 2017. If Congress enacts a new tax provision that offers a substantial credit for capital expenditures made in 2017, then the second firm will have a lower 2017 effective tax rate.

3.4 Corporate Learning and Welfare

This part asks what corporations would learn if they were able to observe each other's tax returns and what the welfare implications of corporate learning would be. The relevant issues are how corporate tax return information would affect tax variation, how the changes in tax variation would affect welfare, and how spending on socially wasteful activities would change. The next part will discuss investor learning—so we may set aside for now any exploration of how tax return information will affect investor forecasts. The arguments generally favor ending corporate tax privacy, but exceptions are discussed in the final subpart.

3.4.1 What might corporations learn?

If corporations could see the universe of corporate returns they would gain valuable information. The two most important ways that they could use this information are: (1) reverse engineer the sheltering and preparation technology of other firms and (2) gain insight into how the IRS picks its battles (Lenter, Slemrod and Shackelford, 2003; Blank, 2014). Both of these would most likely lower tax variation. Firms reverse engineering other firms' sheltering and tax preparation technology would tend to homogenize tax sheltering technology as firms all moved towards having the best tax sheltering technology. Similarly, firms having a better understanding of how the IRS picks its battles would tend to push all firms towards the limit of how much the IRS would tolerate—perhaps with some variation according to the risk preferences of the firm's management.

Blank (2014) studies how corporations might be able to use other corporations' tax returns to improve their tax performance. In particular he suggests that firms could learn about each other's inter-party transactions, see which jurisdiction a firm is tax planning in, and learn about transfer pricing arrangements. Firms might also learn when they have disclosed too much. For example, they might observe a firm in a similar situation not creating a financial reserve for accounting purposes. Lastly, a firm could learn about what the IRS contests and what it does not.

Perhaps more importantly, a corporate tax benchmarking industry would likely emerge. Much the same way corporations are currently measured along customer satisfaction and brand performance metrics, corporations could be measured on their corporate tax performance. Any time a corporation used a new tax shelter, a consultancy could quickly determine whether it would have any value for its corporate clients and if so adapt it for their use. Law firms could examine the universe of corporate tax returns before issuing tax opinions, which presumably would become cheaper as the IRS's interactions with other corporate taxpayers became more public. Activist investors might serve as a catalyst by investing in firms with relatively poor tax performance and increasing shareholder value by upgrading the tax performance—further homogenizing corporate tax technology.

3.4.2 Three models of tax variation

Part III discussed tax variation and in particular noted that there were different sources of tax variation. This subpart postulates three models of tax variation and determines for each whether making corporate returns public would have any effect on tax variation and what the welfare implications of that effect would be. These models are stylized and serve to highlight a concept. Reality is some combination of these models, but for illustrative purposes each is examined independently before they are considered jointly.

Benevolent policy model. As discussed in Part III, there are many reasons to think that optimal policy might place different tax burdens on different firms. For example, taxes could be designed to correct externalities, generate revenue with the smallest possible deadweight loss, or stimulate the economy in a recession. In each of those cases, the optimal tax policy might place a differentiated burden on different corporations.

The tax variation induced by benevolent policy is desirable—indeed the policy is benevolent because its benefits outweigh its costs. Moreover, if benevolent policy were the only source of tax variation, then firms would make no mistakes, have no differentiated tax sheltering technology, and have no power to lobby. Corporations would learn nothing from being able to see other corporations' tax returns, so ending corporate tax privacy would have no effect on tax variation or welfare.

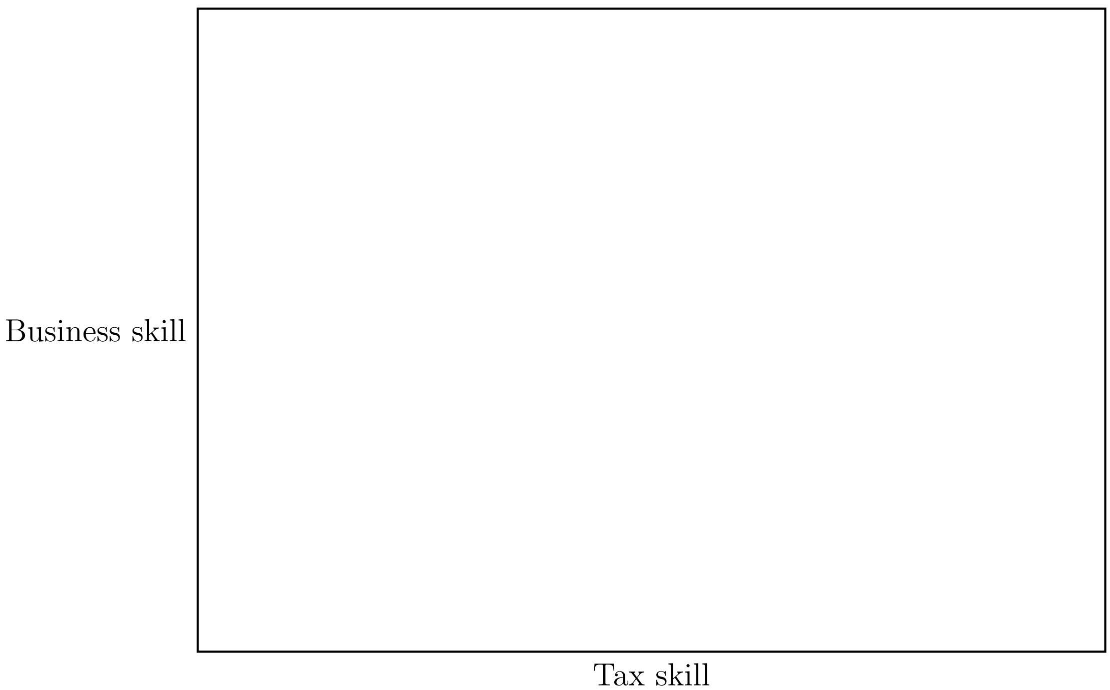
Tax technology variation model. This model posits that firms vary in how good their tax sheltering and tax return preparation technology is. Firms also vary in how productive (and therefore profitable) they are. Investors shift capital around until the marginal after-tax return from every firm is the same. Just as in the GM and Ford example in Part II, total production and income would be highest if investors set marginal pre-tax income equal across all firms, but from that allocation moving investment from a high tax firm to a low tax firm would increase after-tax income. Thus the tax technology variation distorts investment.

To illustrate this more clearly consider the following simplified version of the model. There are several firms, each of which requires a store to go into business. There is an investor who is willing to buy stores for half of the firms if they split their profit with her.

The firms vary in their business skill, and the better firms earn higher profits.[4] The firms also vary in their tax skill—their ability to record their business taxes.[5] Some diligently read the Tax Code, keep careful track of their transactions and business receipts, and judiciously file their returns. Others do none of these things, and most fall between the extremes. If the firms' tax returns were private, other firms would not be able to see them and learn from them—the tax skill that a firm has is the skill it must use.

The figure below captures this model setup. Let every point in the rectangle represent a firm.[6] Firms closer to the top have more business skill and are more profitable. Firms further right have more tax skill and thus have a lower effective tax rate.

Graphical representation of firms



Now assume that returns are private and firms cannot learn from one another. In that case, the investor invests in only the most profitable firms. The pink (light gray) area represents the firms that receive a store.[7] The best firms to invest in are those with high business skill and high tax skill. These firms generate the highest pre-tax income and have the lowest

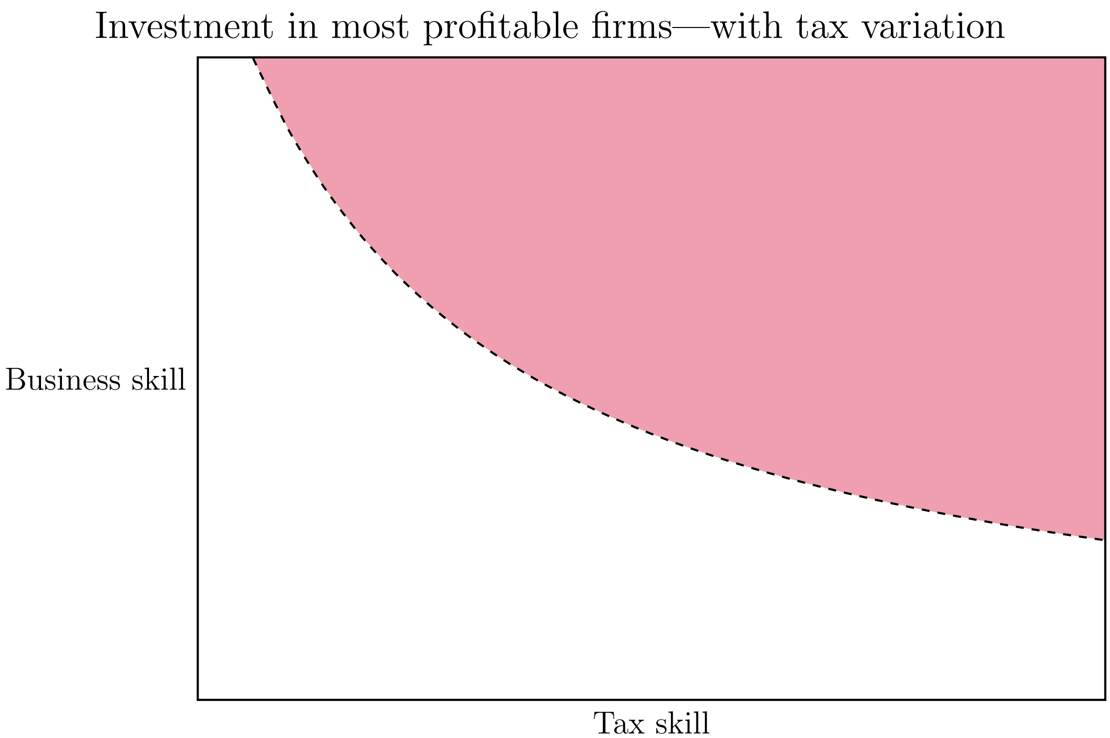
[4] The figures below assume that profit is proportional to business skill.

[5] The figures below assume that (1 - effective tax rate) is proportional to tax skill.

[6] Formally, this analysis assumes that firms are uniformly distributed in the business skill-tax skill space.

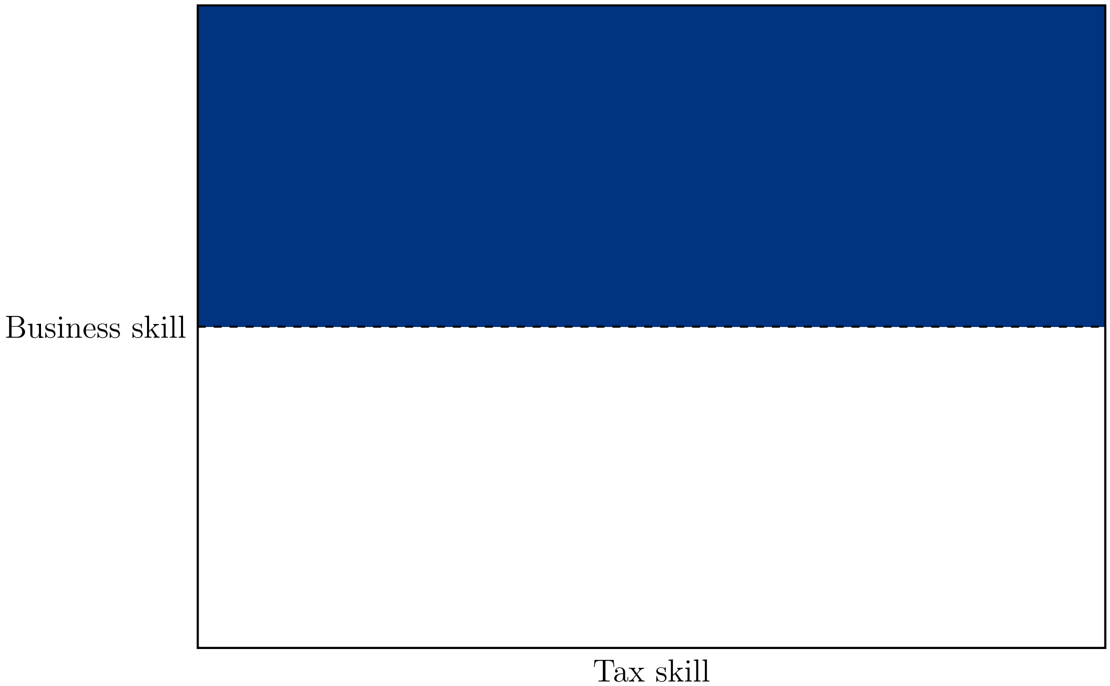
[7] The line bounding the pink (light gray) area is a curved isoprofit line. Let p be profit; t be effective tax rate; and c be a constant. Then the isoprofit line can be derived as follows: $p(1 - t) = c$ implies that $p = c/(1 - t)$. The isoprofit line is curved because a small decrease in tax skill would decrease the profit of a high business skill firm far more than the profit of a low business skill firm.

effective tax rate. Conversely, the worst firms to invest in are those with low business skill and low tax skill. Since the investor can only invest in half of the firms, she will invest in those that are further right and higher up in the rectangle.

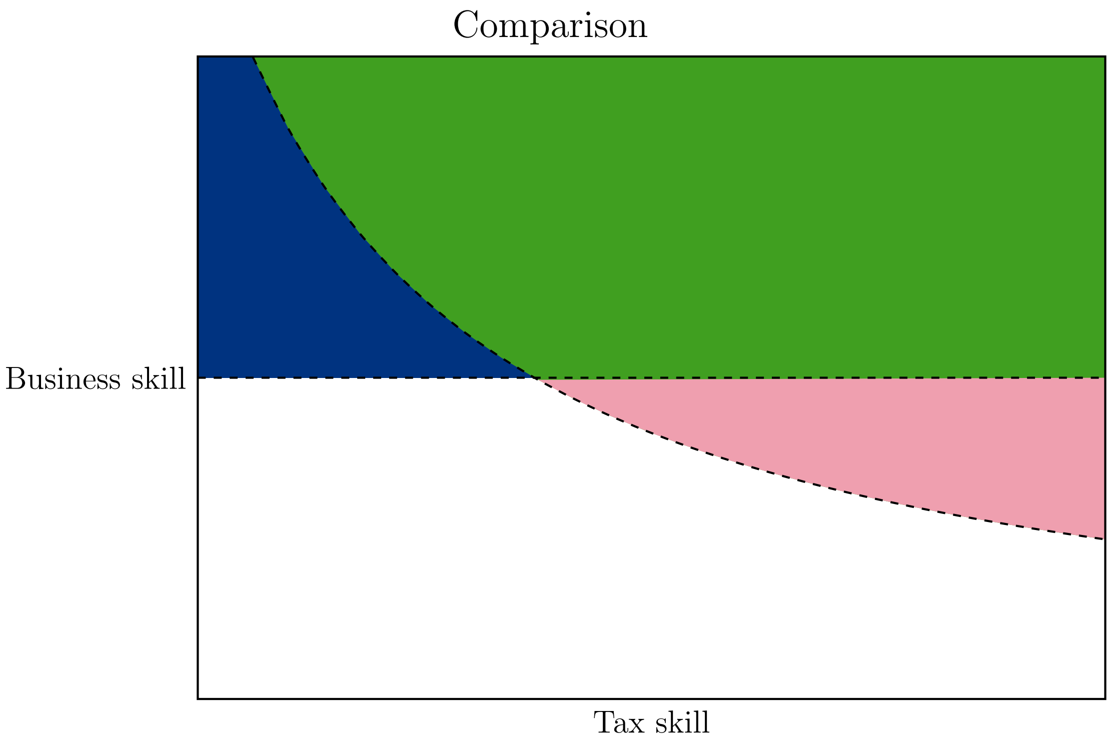


Now assume that returns are public information and firms can learn from one another. Making the extreme assumption that the firms' ability to learn is perfect, the investor will give stores to the firms with the highest business skill, independent of their tax skill because all the firms will have learned the best tax practices. In that case the firms in the blue (dark gray) area in the figure below receive stores.

Investment in most profitable firms—no tax variation



Because every firm has the same tax skill, the investor only cares about investing in the high business skill firms—those that lie in the top half of the rectangle. Figure 4 compares the two scenarios. The green (neutral gray) area represents the firms that will receive a store in both scenarios. The pink (light gray) area represents the firms that will receive a store if returns are private. The blue (dark gray) area represents the firms that will receive a store if returns are public and tax skill is uniform across firms.



The firms in the blue (dark gray) area are more productive than the firms in the pink (light gray) area. Given the constraint that only half the firms will receive investment, the best outcome for society is that the firms with the best business skill receive investment. Because these firms have more business skill, they are able to convert the same inputs into more good and services. However, if investors seek after-tax profits, then they will take into consideration a firm's tax skill when deciding whether it is a worthy investment. This means that some firms that have high tax skill and low business skill will receive investment instead of some of the firms with high business skill and low tax skill. The difference between the business skill of the firms in the blue (dark gray) area and the business skill of the firms in the pink (light gray) is the welfare lost to society because of tax variation, which is undesirable. Ending corporate tax privacy will mitigate the distortion so long as there is some corporate learning from tax returns.

This model can be generalized. First, the investment decision does not have to be binary —it may be that some firms simply receive more investment that others. Second, a firm's business skill and tax skill could be correlated.[8] Third, there could be many different types of business skill. The fundamental point holds nonetheless: if firms vary in their tax skill, this draws investment away from more productive firms towards firms with higher tax skill.

[8] Because the figures assumed a two-dimensional uniform distribution, there implicitly was no correlation between business skill and tax skill.

In this setting, to the extent that it reduces tax skill variation, eliminating corporate tax privacy improves welfare.

Rent-seeking model. Rent-seeking entails using scarce resources in a manner that produces no social value to secure some private benefit.[9] In the rent-seeking model, firms can expend resources to have their tax rate lowered. In particular, firms can hire a tax consultant to help them lower their taxes by increasing tax sheltering.[10] If corporate tax returns are private then firms and consultants cannot learn from each other's work.[11] This means that (1) each firm must pay its own tax consultants and (2) no single tax consultant can look at all the corporate tax returns together and derive the most efficient tax shelters. Consequently, corporate tax privacy causes a large social burden.

In this model, using a tax consultant is an attempt to lower the effective tax rate, which introduces tax variation—this not only distorts investment as in the tax technology variation model, it also funnels scarce resources into redundant and wasteful tax sheltering technology development (Hines, 2004). Several tax consultants would redundantly develop the same tax shelters and write similar tax opinions. Firms need not all have the same preferences over or ability to benefit from rent-seeking. Firm owners or managers might have different abilities to distinguish between good tax consultants and bad tax consultants and different feelings about the duty to pay tax. Variance in rent-seeking would result in tax variation.

Making returns public would disseminate tax-sheltering knowledge. This would lower the cost of tax sheltering which would decrease the distortion of wasteful expenditures on rent-seeking. Similar to the tax technology variation model, making returns public would also limit the misallocation of capital because of tax variation. Ultimately, capital would flow to more productive firms and fewer resources would be wasted on developing tax shelters.

3.4.3 Tax rate variation, corporate tax privacy, and welfare

All three of the above models can explain some of what occurs in reality. Some tax variation is caused by benevolent policy, some is caused by underlying tax technology variation, and some is caused by rent-seeking. This suggests that some tax variation is desirable, some distorts investment, and some is the product of socially wasteful expenditure.

[9] A classic example is several corporations erecting statues of a dictator in the hopes that he will grant them a government contract (Krueger, 1974).

[10] A different version of the model could have political consultants lobby for tax breaks instead of using a tax consultant. If firms were hiring political consultants, then making returns public would only affect tax variation to the extent that public knowledge would stir up activism, either by consumers, investors, or managers.

[11] If the government will raise rates every time a new shelter is discovered, then this model resembles the Prisoner's Dilemma. If firms could agree not to hire tax consultants, they would be better off. However, the optimal strategy is to hire a tax consultant even though it leads to higher rates for everyone.

Recalling the typology developed in Part III, some legal tax variation is beneficial—in particular, the tax variation caused by benevolent policy that corrects externalities, minimizes deadweight loss, or achieves other socially desirable objectives.[12] In contrast, corporate tax variation is not desirable. If firms have different effective tax rates because some make more errors, some have better avoidance schemes, some evade taxes successfully, some make better judgment calls, and some have better luck, then investment allocation will be distorted.

Empirical and anecdotal evidence suggests that firms are not uniformly good at tax preparation and tax sheltering (Rice, 1991). Because investors make investments seeking after-tax profits, differentiated preparation and sheltering technology distorts capital allocation, shifting it away from relatively more productive firms to firms that have relatively better sheltering and preparation technology.

If, however, corporations have access to other corporations' tax returns then they can learn from these returns. This learning process will make tax return preparation and tax sheltering technology more homogenous across firms, making the relative marginal profits of firms depend less on their tax preparation and tax sheltering technologies. Essentially, tax privacy is a barrier preventing firms from having the same tax technology, which allows different firms to play by different tax rules. Instead of the most productive firms receiving investment, those with the most favorable tax technology do.

Making corporate tax returns public will also lower the waste from rent-seeking. Under corporate tax privacy, each firm engaged in rent-seeking has to hire its own consultants, which (bound by confidentiality) are limited in how much they learn from one another—to the extent that tax technology disseminates, it does so slowly. With access to the universe of returns, consultancies would be able to use sophisticated statistical techniques to help firms benchmark tax performance. This too would depress the price of tax consulting services. If returns were public, a handful of consultants could statistically analyze all return data and write software[13] that would help firms find the best tax shelters that were just shy of triggering an IRS dispute. This would be much cheaper than each firm developing shelters independently. Some scholars believe that corporations are near the maximum possible sheltering. [14] If that is the case, it would be better for society to let corporations maintain

[12] The legal tax variation caused by lobbying and inter-governmental competition is harmful. However, this variation is not the focus of this analysis because corporate tax privacy would not have an obvious impact on either lobbying or inter-governmental competition—at least under the theoretical framework developed here.

[13] Some software already exists—for example, GoSystem and Fast Tax. These programs are designed to reduce compliance costs (much like TurboTax), which will lower tax variation. However, a sophisticated program using big data and machine learning would be able to do much more.

[14] Sheppard (2014) quotes Reuven Avi-Yonah, who suggests that we are near the maximum possible corporate aggressiveness.

those shelters without redundant shelter-development expenditures.

3.4.4 Exceptions to the analysis

This subpart explores four scenarios in which the costs from making corporate tax returns public would outweigh benefits and also examines how likely these scenarios are.

Correlation. If the government does not enact benevolent policies, it is possible that firms themselves could generate beneficial tax variation. For example, assume that firms that produce positive externalities are also particularly good at sheltering their taxes. Assuming the government does not appropriately subsidize these firms, they could use their relatively better sheltering technology to achieve the same subsidy a perfectly benevolent government would have provided.

In general, if (1) there is a negative correlation between each firm's optimal tax rate and that firm's tax sheltering technology and (2) the government has not enacted the optimal rate, then corporate tax variation could be welfare improving. If ending corporate tax privacy reduces this beneficial corporate tax variation, then corporate tax privacy is desirable.

While it is true that Congress has not enacted a perfect tax system, there is no evidence that there exists a negative correlation between a firm's socially optimal tax rate and that firm's sheltering ability. Moreover, if the correlation is positive, the case for ending corporate tax privacy is even stronger. In that case, the distorted investment rewards not only less productive firms, but also those with higher externalities.

Increased tax distortions. As noted earlier, eliminating corporate tax privacy will probably decrease government revenue. If the government wishes to recuperate this revenue by increasing other taxes, new economic distortions are likely to emerge. Determining whether eliminating corporate tax privacy is desirable requires a comparison between the distortions under corporate tax privacy and the distortions of the new tax. There are two possible cases in which corporate tax privacy is preferable: (1) the distortion from corporate tax privacy is small or (2) the alternative tax policies have large distortions.

If only a small fraction of firms were actively seeking tax shelters, then the distortion from misallocated investment would be small and the wasteful spending aimed at gaining tax advantages would also be small. In that case, corporate tax privacy might not be a particularly distortionary policy. However, the available evidence suggests that tax planning is widespread and expenditures on tax planning are substantial . Thus the gains from directing investment to the most productive firms and reducing rent-seeking by ending corporate tax privacy likely outweigh the costs (Rice, 1991).

While an analysis of all the possible alternative revenue-raising tax policies is beyond the scope of this paper it seems likely that at least some of them are less distortionary than maintaining corporate tax privacy (Mills, Erickson and Maydew, 1998).

Other barriers to shelter proliferation. A third possibility is that tax preparation and tax sheltering technology are not uniform because of something aside from corporate tax privacy. For example, it might be that large compliance costs prevent smaller companies from using the same tax shelters that large companies use or that tax shelters are particularly well-suited to multinational corporations. In either of those cases firms do not need to learn the sheltering technology, they simply cannot afford it or cannot use it.

While there are surely barriers to using tax shelters beyond corporate tax privacy, eliminating corporate tax privacy is a sufficiently large barrier that removing it would have an impact. If corporate returns were publicly available the cost of tax compliance and procuring tax opinions would fall as firms (and law firms and accountancies) gained a better understanding of how the IRS enforced tax law. Software developed using the data from the universe of corporate returns could also be used to reduce compliance costs. Moreover, even small domestic firms have found ways of using international tax sheltering technology.[15]

It is unlikely that eliminating corporate tax privacy will make tax-sheltering uniform across firms—manager preferences and residual differences in tax skill will still exist. The evidence does, however, suggest that corporate privacy does place a substantial barrier to tax technology proliferation.

3.5 Investor Learning and Welfare

Corporations are not the only ones who would benefit from seeing corporate returns. This part examines what investors would learn from returns, starting from the principles that investors maximize risk-adjusted returns, that doing so requires forecasting future returns, and that investors require information to forecast future returns. It also considers the welfare implications of better forecasts.[16]

3.5.1 Reverse engineering tax liability

If after-tax income were completely determined by pre-tax income, then estimates of pre-tax income would also yield estimates of after-tax income. However, many elements only loosely related (if related at all) to pre-tax income have an effect on tax liability. In addition to knowing a firm's pre-tax income, an investor interesting in determining a firm's tax liability would also need to know how much depreciable capital that firm had and how

[15] Local businesses in a Welsh town used international tax sheltering strategies to minimize their tax liability(Sherwin, Nov. 10, 2015).

[16] Politicians debating the corporate privacy provisions of the 1909 Corporate Excise Tax did consider the usefulness of the information to investors (Kornhauser, 1990).

much that firm had spent on R&D. Neither of these elements is apparent from the income number alone. Thus precise estimates of after-tax income require information about the business and tax decisions firms have made.

Because investors do not have access to tax returns, they turn to other sources for their information on tax liability. For publicly traded companies, the Form 10-K is the most promising source of tax information. Those interested in inferring elements of a corporation's tax return from publicly available documents face challenges (Lisowsky, 2009; Hanlon, 2003). In most cases, financial statements do not state a firm's tax liability. Moreover, the financial statements are substantially different from tax returns and include significantly less information about a firm's tax liability and the elements that comprise it.

Several factors make the 10-K numbers poor proxies for the return numbers. To start with, the two sets of documents are unlikely to be prepared at the same time because the 10-K is due 90 days after the end of a firm's fiscal year whereas the tax return does not need to be filed until eight and a half months have elapsed. This means that when the financial statements are released, the tax return liability need not even have been computed.

More generally, financial statements and tax returns are different because their purposes are fundamentally different. The financial statements apply Generally Accepted Accounting Principles (GAAP) to present information in a way that will help various parties make financial decisions. The corporate tax return records the information necessary to determine tax liability (Hanlon, 2003).

The financial statements do contain some tax information. Most importantly they contain information on GAAP tax expense. Tax expense can be decomposed into two parts. The first is current tax expense, which is intended to capture the amount payable to or refundable by the tax authority. The second part is changes in deferred tax assets and liabilities. For example, if for financial statement purposes an asset is depreciated using the straight-line method and if for tax purposes that same asset is depreciated using an accelerated depreciation method, then the depreciation deduction will be larger (in earlier years) for tax purposes. From this arises a deferred tax liability because the taxes payable are smaller than the accrued tax expense (Hanlon, 2003).

However, current tax expense is different from tax return liability. Michelle Hanlon has researched the most important reasons the numbers differ. The financial statements have different standards for consolidation, are only required to show tax information for continuing operations, and expense stock options as they vest. Furthermore, the provision for uncertain tax positions should accrue probable tax losses and note possible tax losses, but the financial statements are opaque with regard to these reserves. Regression analysis controlling for all of these elements showed that on average for every dollar of current tax expense, tax return liability increases by 70 cents. Compounding the difficulties of inferring tax information from financial statements, firms vary substantially in how much they disclose and how they disclose it (Lisowsky, 2009; Hanlon, 2003).

Firms are required to report other tax information in their financial statements, including a global effective tax rate, net operating losses, permanently reinvested earnings, and penalties and material risks. However, aside from the net operating losses, discretion in disclosure rules makes it difficult to infer exactly what these numbers mean and how they relate to the tax return information (Blank, 2014).

3.5.2 Forecasting future tax liability

Forecasting future tax liability is substantially more difficult than reverse engineering current tax liability. The increased difficulty arises because both a firm's future economic performance and tax environment might change.[17] Without knowing what a firm's current tax liability is, it is difficult to know how different economic circumstances and a different tax environment will affect the tax liability. The possibility of misvaluation is not merely academic: analysts do come up with different valuations because they cannot see corporate tax information (Hanlon, 2003).

As a starting point, a forecaster should know which elements of a firm's taxes are recurring and non-recurring, which are constant and which are likely to change. The SEC does require firms to reconcile their total tax expense to their GAAP income multiplied by their statutory tax rate if the difference is larger than 5%. In some cases, this will yield valuable information. However, reconciliation items need only be listed separately if they account for 5% of the GAAP income times the statutory rate. Thus firms often group items together and provide little description—for example, on some returns credits and deductions are lumped together (Hanlon, 2003).

Evidence suggests that investors, when they can, do take into consideration a firm's exposure to tax risk (Jacob, Wentland and Wentland, 2016). However, the financial statements do not contain much information about the riskiness of tax strategies. Although publicly traded firms are required to keep a tax cushion for uncertain tax positions, many do not and if they do they offer little explanation (Blank, 2014). Hanlon (2003) notes that in many cases it is impossible to tell if a firm's effective tax rate is low because of a credit or a tax shelter.

[17] There is evidence that investors systematically make forecast errors due to book tax differences (Lev and Nissim, 2004; Weber, 2006).

3.5.3 Welfare implications

Information helps reveal which firms earn high returns and which firms earn low returns, which firms are risky and which firms are safe. Investors are better off when they can invest in higher return firms and firms that have risks appropriate to their preferences.[18]

Moreover, if tax rates are the outcome of benevolent policy, then they are designed to encourage investing in some firms and discourage investing in others. For example, an R&D credit is incentivizing investors to fund socially beneficial activities. If, however, investors are unable to observe how much of the credit different firms receive, the intention of the policy will be muffled. In the worst case, investors might mistake a tax credit for a risky tax shelter, which could cause an R&D credit to reduce investment.

The most likely outcome is that increasing investor information would improve welfare. This may not be true if investors misuse the information or find the information too difficult to use. Most investors will, however, benefit from increased information—including all those that have their investments managed (e.g. by purchasing mutual fund shares).

3.6 Conclusion

This paper argues that the information housed in corporate tax returns should be made publicly available. Both corporations and investors could use the information in corporate tax returns to improve welfare.

Public corporate tax return information would homogenize tax return preparation and tax sheltering technology. This would direct investment to firms with relatively higher productivity and away from firms with relatively better preparation and sheltering technology. This would also spare firms the cost of each developing preparation and sheltering technology independently. Firms would be able to implement the tax technology they observed other firms using or the tax technology developed by consultants, who could use the universe of corporate tax returns. Better allocation of investment and less socially wasteful spending would increase productivity, ultimately leading to higher wages, which could increase consumption or leisure.

Tax planning and sheltering are not desirable per se. However, if tax planning and sheltering are already routine, then it is better to lower the barriers that prevent learning about tax planning and sheltering. This prevents firms that are particularly good at tax

[18] Corporate learning, discussed in Section 4, will tend to make firms more homogenous with regards to their tax sheltering and preparation technology. If tax sheltering and preparation technology became perfectly uniform across firms, then firms would not vary in their risk. However, so long as corporate managers have different risk preferences and tax-sheltering technology is imperfect, corporations will have different tax risks. In any event, tax return information will still help investors forecast future expected tax liability.

planning and sheltering from having an economic advantage over other firms and it limits the wasteful, redundant spending on securing tax planning and sheltering advantages.

Investors would also be able to use corporate tax return information to make better forecasts of the returns and risks associated with investing in different firms. Better information would increase investor returns and allow them to choose investments appropriate to their risk appetites. To the extent that corporate taxes are designed to encourage and discourage certain economic activity by changing incentives, better information would make these incentives clearer to investors.

APPENDIX A

Chapter I Supporting Material

A.1 Relationship between b and u

Let $u(x, x_{n+1}) : X \in \mathbb{R}^{n+1} \to \mathbb{R}$ be a strictly increasing, strictly concave utility function over the choice set X of a representative agent. Let $p(x, x_{n+1}) : \mathbb{R}^{n+1} \to \mathbb{R}$ be a strictly decreasing, weakly convex function, such that $p = 0$ defines a production possibility frontier. Under the implicit mapping theorem there exists a function f such that for all (x, x_{n+1}) such that $p(x, x_{n+1}) = 0$, $x_{n+1} = f(x)$. $p(x, f(x)) = 0$ and $f_i = -p_i/p_{n+1}$.

Proposition A.1.1. *Define* $b(x) = u(x, f(x))$. *Then*

(i) x^* *maxmizes* b *if and only if* $(x^*, f(x^*))$ *maximizes* u *subject to* $p = 0$*;*

(ii) $\max_x\{b(x)\} = \max_{x,x_{n+1}}\{u(x, x_{n+1})$ *s.t.* $p(x, x_{n+1}) = 0\}$*; and*

(iii) b *is strictly concave.*

Proof. (i) The first order conditions of the utility maximization problem are $\forall_{i\in\{1,\ldots,n\}}$: $\frac{u_i}{u_{n+1}} = \frac{p_i}{p_{n+1}}$. The first order conditions of the private net benefit function are $b_i = 0 \implies u_i + u_{n+1}f_i = u_i + u_{n+1}(-p_i/p_{n+1}) = 0 \implies \frac{u_i}{u_{n+1}} = \frac{p_i}{p_{n+1}}$. (ii) follows.
(iii) First, note that weakly convex p implies weakly concave f.

$$
\begin{aligned}
\forall(x, x_{n+1}), (y, y_{n+1}) \text{ s.t. } p(x, x_{n+1}) &= p(y, y_{n+1}) = 0 \\
p(y, y_{n+1}) - p(x, x_{n+1}) &\leq \sum_{i=1}^{n+1} p_i(y_i - x_i) \implies \\
0 &\leq \sum_{i=1}^{n+1} p_i(y_i - x_i) \implies \\
-p_{n+1}(y_{n+1} - x_{n+1}) &\leq \sum_{i=1}^{n} p_i(y_i - x_i) \implies \\
y_{n+1} - x_{n+1} &\geq \sum_{i=1}^{n} (-p_i/p_{n+1})(y_i - x_i) \implies \\
f(y) - f(x) &\geq \sum_{i=1}^{n} (-p_i/p_{n+1})(y_i - x_i) \implies \\
f(y) - f(x) &\geq \sum_{i=1}^{n} f_i(y_i - x_i)
\end{aligned}
$$

$$
\begin{aligned}
&\forall \theta \in [0, 1] \\
b\,(\theta x + [1-\theta]y) &= u(\theta x + [1-\theta]y, f(\theta x + [1-\theta]y)) \\
&\geq u(\theta x + [1-\theta]y, \theta f(x) + [1-\theta]f(y)) \\
&= u(\theta x + [1-\theta]y, \theta x_{n+1} + [1-\theta]y_{n+1}) \\
&> \theta u(x, x_{n+1}) + [1-\theta]u(y, y_{n+1}) \\
&= \theta b(x) + [1-\theta]b(y)
\end{aligned}
$$

□

This shows that the private net benefit problem generalizes the constrained utility problem. There are two points worth noting. First, there is no tax on x_{n+1} when the private net benefit function is used. Second, using the b function removes prices from the problem, which means that the tax must be in the same units as the objective function.

A.2 The details of the output tax problem

Strictly increasing, weakly concave q maps activities to output. Strictly increasing, strictly concave v maps output to private benefit. Strictly increasing, weakly convex g maps the activities to private cost. $b(x) = v(q(x)) - g(x)$

Proposition A.2.1. *The best response function $x(\tau)$ exists and is continuously differentiable.*

Proof. The private market's problem is:

$$\max_x v(q(x)) - g(x) - \tau q(x)$$

The first order condition is $v'(q(x))q'(x) - g'(x) - \tau q'(x) = 0$. Taking the derivative of first order condition with respect to x

$$\begin{aligned} &v'(q(x))q''(x) + (q'(x))^\top v''(q(x))q'(x) - g''(x) - \tau q''(x) \\ =&\underbrace{(v'(q(x)) - \tau)}_{\text{FOC} \implies +} \underbrace{q''(x)}_{\text{neg. def.}} + \underbrace{v''(q(x))}_{-} \underbrace{(q'(x))^\top q'(x)}_{\text{pos. sem. def.}} - \underbrace{g''(x)}_{\text{pos. sem. def.}} \end{aligned}$$

Since a positive semidefinite matrix multiplied by a negative scalar is a negative semidefinite matrix and since the sum of a negative definite matrix and to semidefinite matrices is a negative definite matrix, the implicit mapping theorem implies that $x(\tau)$ exists and is continuously differentiable. As an aside, $b = v(q) - g$ is concave.

□

Proposition A.2.2. $q'(x(\tau))x'(\tau) < 0$

Proof. Take the derivative of first order condition with respect to τ. Note that the first order condition must still hold, so the derivative with respect to τ will equal 0.

$$\begin{aligned} v'q''x' + (q')^\top v''q'x' - g''x' - \tau q''x' - (q')^\top = 0 &\implies \\ (x')^\top v'q''x' + (x')^\top (q')^\top v''q'x' - (x')^\top g''x' - (x')^\top \tau q''x' = (q')^\top &\implies \\ \underbrace{(v' - \tau)}_{+} \underbrace{(x')^\top q''x'}_{-} + \underbrace{(q'x')^2}_{+} \underbrace{v''}_{-} - \underbrace{(x')^\top g''x'}_{+} = (q'x')^\top &\implies \\ (q'x')^\top < 0 & \end{aligned}$$

$\square$

A.3 Comparing activity taxes

Consider two arbitrary activity tax vectors, v and w. Assuming no administrative cost, the change in welfare of moving from v to w is:

$$\begin{aligned}\Delta b - e^\top \Delta x &= b(x(w)) - b(x(v)) - e^\top [x(w) - x(v)] \\ &= b(x(\gamma(1))) - b(x(\gamma(0))) - e^\top [x(\gamma(1)) - x(\gamma(0))]\end{aligned}$$

where $\gamma : \mathbb{R} \to \mathbb{R}^n,\ \gamma(r) = v + r(w - v)$. Then by the fundamental theorem of calculus:

$$= \int_0^1 [b'(x(\gamma(r))) - e^\top]\ x'(\gamma(r))\ \gamma'(r)\ dr$$

Recalling that $b'(x(t)) = t^\top$:

$$= \int_0^1 (\gamma(r) - e)^\top\ x'(\gamma(r))\ \gamma'(r)\ dr$$

Integrating by parts:

$$\begin{aligned}&= (\gamma(r) - e)^\top\ x(\gamma(r))\Big|_{r=0}^1 \ - \int_0^1 \gamma'(r)^\top\ x(\gamma(r))\ dr \\ &= (w - e)^\top x(w) - \ (v - e)^\top x(v) \ - \ (w - v)^\top \int_0^1 x(\gamma(r))\ dr\end{aligned}$$

Assuming $\int_0^1 x(\gamma(r))\ dr \approx \frac{1}{2}(x(w) + x(v))$, which is analogous to Harberger (1964)'s approximation:

$$\begin{aligned}&= (w - e)^\top x(w) - \ (v - e)^\top x(v) \ - \ \frac{1}{2}(w - v)^\top (x(w) + x(v)) \\ &= \left[(w - e) - \ \frac{1}{2}(w - v)\right]^\top x(w) - \ \left[(v - e) \ + \ \frac{1}{2}(w - v)\right]^\top x(v) \\ &= \frac{1}{2}(w + v - 2e)^\top (x(w) - x(v))\end{aligned}$$

If $v = e$ (i.e. the social planner shifts away from the optimal activity tax), then:

$$\Delta b - e^\top \Delta x = \int_0^1 r(w - e)^\top\ x'(\gamma(r))\ (w - e)\ dr \tag{A.1}$$

Note that the integral is the sum of negative definite matrices and that the expression is a quadratic form. Thus the optimal activity tax is strictly superior to any other activity tax. Note that the welfare lost increases with the square of the uncorrected externality and increases with the weighted average size of $x'(t)$ over the set $[e, w]$. $x'(t)$ is weighed by r because close to the optimum the marginal benefit and marginal social harm are equal but diverge more the further w is from e—for the same reason dead weight loss approximations are a triangle.

As an aside, this appendix generalizes the *Harberger Triangle* to a setting with more than one taxed good and an initial set of taxes.[1] Assuming there are no externalities, the approximate change in welfare from moving to one tax vector to another is:

$$\Delta b \approx \frac{1}{2}(w + v)^{\top}(x(w) - x(v)) \tag{A.2}$$

This equation corresponds to the *Harberger Trapezoid* associated with a change from an existing set of taxes to another set of taxes. Setting $v = 0$ simplifies this equation to the Harberger Triangle associated with a change from no taxes to a new set of taxes, in a setting with more than one good and more than one tax:

$$\Delta b \approx \frac{1}{2}w^{\top}(x(w) - x(0)) \tag{A.3}$$

[1] See generally Auerbach (1985) and Auerbach and Hines (2002).

APPENDIX B

Chapter II Supporting Material

B.1 Estimate of the Cost

Let a_i be the arrests for crime i, x_j be the quantity of crime j, and $C(\mathbf{x}(\mathbf{a}), \mathbf{q}(\mathbf{a}))$ be the total cost of crime. We define $q_i(a_i) = a_i \ \forall i$ for notational convenience. Each index crime j is assumed to be a function of arrests for each index crime. We assume that $\frac{\partial C}{\partial q_i}$ and $\frac{\partial C}{\partial x_i}$ are constant. Thus the total cost of crime is:

$$\sum_i \frac{\partial C}{\partial q_i} a_i + \sum_j \frac{\partial C}{\partial x_j} x_j(a_1, ..., a_n) = \sum_i A_i a_i + \sum_j B_j x_j(a_1, ..., a_n)$$

where $A_i = \frac{\partial C}{\partial q_i}$ and $B_j = \frac{\partial C}{\partial x_j}$. The total differential of $x_j(\mathbf{a})$ is:

$$dx_j = \sum_i \frac{\partial x_j}{\partial a_i} da_i = \sum_i \varepsilon(x_j, a_i) \frac{x_j}{a_i} da_i \implies \frac{dx_j}{x_j} = \sum_i \varepsilon(x_j, a_i) \frac{da_i}{a_i}$$

where $\varepsilon(y, z)$ is the elasticity of y with respect to z. Assuming that the elasticities are constant:

$$\int \frac{1}{x_j} dx_j = \sum_i \left[\varepsilon(x_j, a_i) \int \frac{1}{a_i} da_i \right] \implies \ln(x_j) = \sum_i \varepsilon(x_j, a_i) \ln(a_i) + k_j$$

where k_j is a constant of integration. We have:

$$x_j = \exp\left[\sum_i \varepsilon(x_j, a_i) \ln(a_i) + k_j \right] = e^{k_j} \prod_i \exp\left[\varepsilon(x_j, a_i) \ln(a_i)\right] = e^{k_j} \prod_i a_i^{\varepsilon(x_j, a_i)}$$

We then return to the cost function as a function of $\mathbf{a}$ only:

$$C(\mathbf{a}) = \sum_i A_i a_i + \sum_j \left[B_j e^{k_j} \prod_i a_i^{\varepsilon(x_j, a_i)} \right]$$

Our aim is:

$$\min_{a_i \geq 0} C(\mathbf{a}) \text{ such that } \sum_i A_i a_i = \bar{A}$$

where $\bar{A}$ is the total cost of arrest at the current level of arrests.

Solving this problem for index crimes requires estimates of A_i, B_j, and $\varepsilon(x_j, a_i)$. McCollister et. al. provides B_j. We use the reported information in McCollister et. al., Levitt, and the UCR to find A_i and $\varepsilon(x_j, a_i)$.

Levitt reports elasticities of crimes with respect to arrest rates for own price elasticities and average arrest rates for cross price elasticities. For the own price elasticities, the elasticity with respect to arrests, $\varepsilon(x_i, a_i)$, is the same as Levitt's reported elasticity with respect to arrest rates, $\varepsilon(x_i, \frac{a_i}{x_i})$,:

$$\varepsilon(x_i, \frac{a_i}{x_i}) = \frac{\partial x_i}{\partial (a_i/x_i)} \frac{a_i/x_i}{x_i} \implies$$

$$\varepsilon(x_i, a_i) = \frac{\partial x_i}{\partial a_i} \frac{a_i}{x_i} = \frac{\partial x_i}{\partial (a_i/x_i)} \frac{\partial a_i/x_i}{\partial a_i} \frac{a_i}{x_i} = \varepsilon(x_i, \frac{a_i}{x_i}) \frac{x_i}{a_i/x_i} \frac{1}{x_i} \frac{a_i}{x_i} = \varepsilon(x_i, \frac{a_i}{x_i})$$

For cross price elasticities we must modify Levitt's estimates. To convert Levitt's elasticities to $\varepsilon(x_j, a_i)$, we use the 2010 nationwide counts of crime and arrests from the UCR. Levitt reports the elasticity with respect to the arrest rate for substitute and non-substitute crimes, where violent crimes are assumed to be substitutes with each other and non-substitutes with pecuniary crimes and vice-versa. The equation for computing the cross price elasticity is the same for substitutes and non-substitutes. For example, if Levitt's categories imply that crime j is not a substitute for crime i, then using k as an index for all of the crimes that Levitt categorizes as non-substitute crimes including i:

$$\varepsilon(x_j, \frac{\sum a_k}{\sum x_k}) = \frac{\partial x_j}{\partial (\frac{\sum a_k}{\sum x_k})} \frac{(\frac{\sum a_k}{\sum x_k})}{x_i} \implies$$

$$\varepsilon(x_j, a_i) = \frac{\partial x_i}{\partial(\frac{\sum a_k}{\sum x_k})} \frac{\partial(\frac{\sum a_k}{\sum x_k})}{\partial a_i} \frac{a_i}{x_i} = \varepsilon(x_j, \frac{\sum a_k}{\sum x_k}) \frac{x_i}{(\frac{\sum a_k}{\sum x_k})} \frac{1}{\sum x_k} \frac{a_i}{x_i} = \varepsilon(x_j, \frac{\sum a_k}{\sum x_k}) \frac{a_i}{\sum a_k}$$

If Levitt considers crime j a substitute for crime i, then using h as an index for all of the crimes Levitt considers substitutes (j is not included in this set of crimes):

$$\varepsilon(x_j, \frac{\sum a_h}{\sum x_h}) = \frac{\partial x_j}{\partial(\frac{\sum a_h}{\sum x_h})} \frac{(\frac{\sum a_h}{\sum x_h})}{x_i} \implies$$

$$\varepsilon(x_j, a_i) = \frac{\partial x_i}{\partial(\frac{\sum a_h}{\sum x_h})} \frac{\partial(\frac{\sum a_h}{\sum x_h})}{\partial a_i} \frac{a_i}{x_i} = \varepsilon(x_j, \frac{\sum a_h}{\sum x_h}) \frac{x_i}{(\frac{\sum a_h}{\sum x_h})} \frac{1}{\sum x_h} \frac{a_i}{x_i} = \varepsilon(x_j, \frac{\sum a_h}{\sum x_h}) \frac{a_i}{\sum a_h}$$

Since Levitt reports a small positive own price elasticity for murder, this approach would imply that additional arrests for murder reduce costs, even in the absence of an effect on crime levels. To resolve this we replace the entry for murder's own-price elasticity with a small negative value.

The Slutsky matrix is made up of these elasticities of demand for each crime with respect to the number of each type of arrest. The Slutsky matrix is not symmetric which may be because there are income effects and heterogenous agents.

We can make the Slutsky matrix symmetric by averaging each pair of off diagonal terms, computing $\frac{\varepsilon(x_j,a_i)+\varepsilon(x_i,a_j)}{2}$ and using that value for both $\varepsilon(x_j, a_i)$ and $\varepsilon(x_j, a_i)$. Leaving the Slutsky matrix asymmetric leads to similar results.

A_i is a measure of the costliness of arrests of type i computed as follows. If the arrest rates are set independently in order to control each crime separately, the own-price-elasticities and the victim harms imply a marginal cost A_i of additional arrests of type i. This is equivalent to assuming that current punishments correspond to the naive punishments and computing the cost per arrest which satisfies the naive first order conditions. Equating the marginal cost of additional arrests for crime i with the marginal decrease in harm to victims associated with crime i and rewriting the expression in terms of the information from McCollister et. al., Levitt, and the UCR:

$$A_i = \frac{\partial C}{\partial x_i}\frac{\partial x_i}{\partial a_i} \implies A_i = B_i \varepsilon(x_i, \frac{a_i}{x_i})\frac{x_i}{a_i}$$

The vectors of arrests which satisfy the constraint $\sum_i A_i a_i = \bar{A}$ can be constructed by creating random vectors in the null space of $\mathbf{A}$ and adding them to any vector which satisfies the constraint. By constructing many random combinations of arrests which satisfy the constraint, we can simply check which of those combinations of arrests has the lowest cost. We restrict the search to a region around a starting value, initially set at 2010 arrest rates, and repeat the process starting with the combination of arrests associated with the lowest cost from the previous iteration.

B.2 Tables

Table B.1Previous empirical work

Marvell and Moody (1994)	Index crimes	Total prison population	Murder -0.065; Agg Assault 0.056; Robbery -0.260; Rape -0.113; Burglary -0.253; Larceny -0.138; Auto theft -0.200
Marvell and Moody (1996)	Index crimes	Number of police officers	Murder -0.36; Agg Assault -0.35; Robbery -0.63; Rape -0.20; Burglary -0.33; Larceny -0.22; Auto theft -0.85
Evans and Owens (2007)	Index crimes	Number of police officers	Elasticities: Murder -0.84; Rape -0.42; Agg Assault -0.96; Robbery -1.34; Burglary -0.59; Larceny -0.08; Auto theft -0.85
Levitt (1997)	Index crimes	Number of police officers	2SLS results Murder $\in [-1.18, -3.05]$; Rape $\in [0.67, -0.27]$; Agg Assault $\in [-0.36, -1.09]$; Robbery $\in [-0.38, -1.20]$; Burglary $\in [-0.05, -0.58]$; Larceny $\in [0.26, -0.43]$; Auto theft $\in [0.14, -0.61]$
Corman and Mocan (2000)	Murder, robbery, burglary, auto theft	Arrests; police	Elasticity with respect to arrests: Murder -0.34; Robbery -0.94; Burglary -0.36; Auto Theft -0.40
Shepherd (2002)	Index crimes	Truth In Sentencing legislation (increases punishment for violent felonies)	Murder: -1.178; Agg Assault -44.809; Robbery -39.615; Rape -4.226; Burglary 174.721; Larceny -89.486; Auto theft 70.252

Study	Crimes studied	'Prices'	Result
Benson et al. (1992)	Aggregate Property	Probability of arrest	-0.826
Levitt (1998a)	Aggregate violent; aggregate property	Differences in juvenile and adult punitiveness	Violent -0.121 Property -0.050
Kessler and Levitt (1999)	Aggregate crime	Sentence enhancement legislation in California	After legislation, crime rates for crimes eligible for sentence enhancement diverge from ineligible crimes in California, relative to the national rates
Shepherd (2004)	Types of murder	Probability of sentence; probability of execution	Elasticities for seven subcategories of murder by intimates: -0.04; Acquaintance -0.012; Stranger -0.002; Crime of Passion -0.04; Felonies -0.001; White -0.017; African American -0.02
Ehrlich (1975)	Murder	Probability of arrest, conditional probability of conviction, conditional probability of execution	Variety of specifications, reporting point estimates: $P_a \in [-1.182, -2.225]$; $P_{c\|a} \in [-0.374, -0.850]$; $P_{e\|c} \in [-0.039, -0.068]$
Dezhbakhsh, Rubin and Shepherd (2003)	Murder	Probability of arrest, conditional probability of conviction, conditional probability of execution	$P_a \in [-2.184, -10.096]$; $P_{c\|a} \in [-3.597, -47.661]$; $P_{e\|c} \in [-2.715, -7.396]$

Study	Crimes studied	'Prices'	Result
Durlauf, Navarro and Rivers (2010)	Murder	Probability of Arrest, Death Sentence and/or Execution	Net lives saved can vary greatly based on specification
Duggan (2001)	Gun and nongun homicide; all index crimes	One year lagged gun ownership (may be a proxy for victim precaution or a proxy for ease of committing violent crime)	Gun Homicide: 0.306; 0.223; nongun homicide: 0.020; 0.040; homicide: 0.180; 0.210; Agg Assault -0.007; -0.013; Robbery -0.016; 0.069; Rape -0.052; -0.092; Burglary -0.002; 0.094; Larceny 0.081; 0.032; Auto theft 0.043; 0.019
Lott and Mustard (1997)	Index crimes	'shall issue' laws (gun ownership - may be a proxy for victim precaution or a proxy for ease of committing violent crime)	Murder: -0.049; Agg Assault -0.0701; Robbery -0.0221; Rape -0.0527; Burglary 0.00048; Larceny 0.03342; Auto theft 0.0714
Kuziemko and Levitt (2004)	Aggregate crime	Share of prisoners incarcerated for violent, pecuniary, and drug crime	Cannot reject equal impact at the margin of incarcerating different types of prisoners

Table B.**2**The impact of a small change in each arrest rate

	Changes in each type of offense due to a 1% change in each arrest rate							Total harm	Harm reduction from
1% increase in	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto Theft	reduction	cross crime effects
Murder arrests	-1.47	-0.18	5.83	-14.64	-5.28	-36.51	-8.25	$14m	$1.6m
Rape arrests	-0.71	-92.44	10.45	-26.25	-9.47	-65.47	-14.80	$31m	$9.2m
Robbery arrests	-3.94	-1.81	-1225.38	-146.75	-52.93	-366.01	-82.74	$106m	$54m
Assault arrests	-14.34	-6.57	212.55	-1305.68	-192.54	-1331.35	-300.97	$270m	$131m
Burglary arrests	-1.05	-25.22	-102.19	-27.75	-6765.59	2388.87	-409.06	$62m	$19m
Larceny arrests	-4.59	-110.65	-448.38	-121.77	-1272.88	-17869.25	-1794.84	$191m	$127m
Auto theft arrests	-0.26	-6.22	-25.21	-6.85	-71.57	589.34	-547.28	$9.9m	$4.0m

BIBLIOGRAPHY

Aizer, Anna, and Joseph J Doyle. 2015. "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges." *The Quarterly Journal of Economics*, qjv003.

Anderson, David A. 1999. "The Aggregate Burden of Crime." *The Journal of Law and Economics*, 42(2): 611–642.

Auerbach, Alan J. 1985. "The theory of excess burden and optimal taxation." *Handbook of Public Economics*, 1: 61–127.

Auerbach, Alan J, and James R Hines. 2002. "Taxation and economic efficiency." *Handbook of Public Economics*, 3: 1347–1421.

Auger, Pat, Paul Burke, Timothy M Devinney, and Jordan J Louviere. 2003. "What will consumers pay for social product features?" *Journal of business ethics*, 42(3): 281–304.

Beck, Allen J, and Bernard E Shipley. 1989. *Recidivism of prisoners released in 1983.* US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics Washington, DC.

Becker, Gary. 1968. "Punishment: An Economic Approach." *Journal of Political Economy*, 169: 169–217.

Bell, Brian, and Stephen Machin. 2013. "Immigration and crime." *A. Constant and K. Zimmermann, International Handbook on the Economics of Migration, Edward Elgar Publishing*, 353–372.

Benson, Bruce L, David W Rasmussen, and Iljoong Kim. 1998. "Deterrence and public policy: trade-offs in the allocation of police resources." *International Review of Law and Economics*, 18(1): 77–100.

Benson, Bruce L, Iljoong Kim, David W Rasmussen, and Thomas W Zhehlke. 1992. "Is property crime caused by drug use or by drug enforcement policy?" *Applied Economics*, 24(7): 679–692.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment." *The American Economic Review*, 105(11): 3489–3529.

Blank, Joshua D. 2014. "Reconsidering Corporate Tax Privacy." *New York University Journal of Law & Business*, 11: 31.

Bloomberg. Jan. 2015. "Top Tax & Accounting Mistakes That Cost Companies Millions."

Blumenstein, Rebecca, Dennis K. Berman, and Evan Perez. May 2, 2003. "After Inflating Their Income, Companies Want IRS Refunds." *Wall Street Journal*.

Bovenberg, A Lans, and Frederick van der Ploeg. 1994. "Environmental policy, public finance and the labour market in a second-best world." *Journal of Public Economics*, 55(3): 349–390.

Bovenberg, A Lans, and Lawrence H Goulder. 1996. "Optimal environmental taxation in the presence of other taxes: general-equilibrium analyses." *The American Economic Review*, 985–1000.

Bovenberg, A Lans, and Ruud A De Mooij. 1994. "Environmental levies and distortionary taxation." *The American Economic Review*, 1085–1089.

Corlett, Wilfred J, and Douglas C Hague. 1953. "Complementarity and the excess burden of taxation." *The Review of Economic Studies*, 21–30.

Corman, Hope, and H Naci Mocan. 2000. "A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City." *American Economic Review*, 584–604.

Cremer, Helmuth, and Firouz Gahvari. 1993. "Tax evasion and optimal commodity taxation." *Journal of Public Economics*, 50(2): 261–275.

Cremer, Helmuth, and Firouz Gahvari. 2001. "Second-best taxation of emissions and polluting goods." *Journal of Public Economics*, 80(2): 169–197.

Cremer, Helmuth, Firouz Gahvari, and Norbert Ladoux. 1998. "Externalities and optimal taxation." *Journal of Public Economics*, 70(3): 343–364.

Dezhbakhsh, Hashem, Paul H Rubin, and Joanna M Shepherd. 2003. "Does capital punishment have a deterrent effect? New evidence from postmoratorium panel data." *American Law and Economics Review*, 5(2): 344–376.

Duggan, Mark. 2001. "More Guns, More Crime." *Journal of Political Economy*, 109(5).

Durlauf, Steven N, Salvador Navarro, and David A Rivers. 2010. "Understanding aggregate crime regressions." *Journal of Econometrics*, 158(2): 306–317.

Egan, Paul. Jul. 20, 2015. "Snyder Signs Bill Ending Incentives for Film Industry." *Detroit Free Press*.

Ehrlich, Isaac. 1973. "Participation in illegitimate activities: A theoretical and empirical investigation." *The Journal of Political Economy*, 521–565.

Ehrlich, Isaac. 1975. "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *The American Economic Review*, 65(3): 397–417.

Ernst & Young. Feb. 2011. "Economic and Fiscal Impacts of the Michigan Film Tax Credit."

Evans, William N, and Emily G Owens. 2007. "COPS and Crime." *Journal of Public Economics*, 91(1): 181–201.

Fullerton, Don. 1997. "Environmental levies and distortionary taxation: comment." *The American Economic Review*, 245–251.

Fullerton, Don, and Ann Wolverton. 2005. "The two-part instrument in a second-best world." *Journal of Public Economics*, 89(9): 1961–1975.

Fullerton, Don, and Sarah E West. 2002. "Can taxes on cars and on gasoline mimic an unavailable tax on emissions?" *Journal of Environmental Economics and Management*, 43(1): 135–157.

Gahvari, Firouz. 2014. "Second-best pigouvian taxation: a clarification." *Environmental and Resource Economics*, 59(4): 525–535.

Hakim, Simon, Uriel Spiegel, and Jimmy Weinblatt. 1984. "Substitution, size effects, and the composition of property crime." *Social Science Quarterly*, 65(3): 719–34.

Hanlon, Michelle. 2003. "What can we infer about a firm's taxable income from its financial statements?" *National Tax Journal*, 831–863.

Hanlon, Michelle, and Joel Slemrod. 2009. "What does tax aggressiveness signal? Evidence from stock price reactions to news about tax shelter involvement." *Journal of Public Economics*, 93(1): 126–141.

Harberger, Arnold C. 1964. "The measurement of waste." *American Economic Review*, 58–76.

Heineke, John. 1978. "Economic models of criminal behavior: An overview." *North-Holland Publishing Company.*

Hines, James R. 2004. "On the timeliness of tax reform." *Journal of Public Economics*, 88(5): 1043–1059.

Hines, James R, and Richard H Thaler. 1995. "Anomalies: The flypaper effect." *The Journal of Economic Perspectives*, 9(4): 217–226.

Hjelmgaard, Kim, and Kevin McCoy. Aug. 6, 2014. "Walgreen's Stock Smacked after Tax Inversion out." *USA TODAY.*

Internal Revenue Service. 2016. "Budget in Brief FY2017."

Internal Revenue Service. Dec. 2011. "Tax Gap "Map" Tax Year 2006."

Internal Revenue Service. Oct. 2, 2015. "Examples of Corporate Fraud Investigations—Fiscal Year 2015."

Jacob, Martin, Kelly Wentland, and Scott Wentland. 2016. "Real Effects of Tax Uncertainty: Evidence from Firm Capital Investments." *SSRN Working Paper.*

Jacobs, Bas, and Ruud A de Mooij. 2015. "Pigou meets Mirrlees: On the irrelevance of tax distortions for the second-best Pigouvian tax." *Journal of Environmental Economics and Management*, 71: 90–108.

Jacobsen, Mark R., Christopher R. Knittel, James M. Sallee, and Arthur A. van Benthem. 2016. "Sufficient Statistics for Imperfect Externality-Correcting Policies." *NBER Working Paper No. 22063*, March.

Kaplow, Louis. 1990*a*. "A note on the optimal use of nonmonetary sanctions." *Journal of Public Economics*, 42(2): 245–247.

Kaplow, Louis. 1990*b*. "Optimal taxation with costly enforcement and evasion." *Journal of Public Economics*, 43(2): 221–236.

Kaplow, Louis. 2012. "Optimal control of externalities in the presence of income taxation." *International Economic Review*, 53(2): 487–509.

Kessler, Daniel, and Steven D Levitt. 1999. "Using Sentence Enhancements to Distinguish Between Deterrence and Incapacitation." *Journal of Law and Economics*, 42(1): 343–363.

Kocieniewski, David. Mar. 24, 2011. "GE's strategies let it avoid taxes altogether." *New York Times.*

Kopczuk, Wojciech. 2003. "A Note on optimal taxation in the presence of externalities." *Economics Letters*, 81–86.

Kornhauser, Marjorie. 1990. "Corporate Regulation and the Origin of the Corporate Income Tax." *Indiana Law Journal*, 66: 53.

Kornhauser, Marjorie E. 2005. "Doing the Full Monty: Will Publicizing Tax Information Increase Compliance?" *Canadian Journal of Law & Jurisprudence*, 18(1): 95–117.

Krueger, Anne O. 1974. "The political economy of the rent-seeking society." *The American economic review*, 64(3): 291–303.

Kuziemko, Ilyana, and Steven D Levitt. 2004. "An empirical analysis of imprisoning drug offenders." *Journal of Public Economics*, 88(9): 2043–2066.

Leiser, Stephanie. 2014. "Essays on State Business Tax Incentives and Policy Diffusion." PhD diss. University of Washington.

Lenter, David, Joel Slemrod, and Douglas Shackelford. 2003. "Public disclosure of corporate tax return information: Accounting, economics, and legal perspectives." *National Tax Journal*, 803–830.

Lev, Baruch, and Doron Nissim. 2004. "Taxable income, future earnings, and equity values." *The Accounting Review*, 79(4): 1039–1074.

Levitt, Steven D. 1997. "Using electoral cycles in police hiring to estimate the effect of police on crime." *The American Economic Review*, 270–290.

Levitt, Steven D. 1998*a*. "Juvenile Crime and Punishment." *Journal of Political Economy*, 106(6): 1156–1185.

Levitt, Steven D. 1998*b*. "Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error?" *Economic inquiry*, 36(3): 353–372.

Lipsey, Richard G., and Kelvin Lancaster. 1956. "The general theory of the second best." *The Review of Economic Studies*, 24(1): 11–32.

Lisowsky, Petro. 2009. "Inferring US tax liability from financial statement information." *Journal of the American Taxation Association*, 31(1): 29–63.

Lott, Jr, John R, and David B Mustard. 1997. "Crime, deterrence, and right-to-carry concealed handguns." *The Journal of Legal Studies*, 26(1): 1–68.

Lott, Jr., John R, and William M Landes. 2000. "Multiple victim public shootings." *University of Chicago Law and Economics Working Paper*.

Marvell, Thomas B, and Carlisle E Moody. 1996. "Specification Problems, Police Levels, And Crime Rates." *Criminology*, 34(4): 609–646.

Marvell, Thomas B, and Carlisle E Moody, Jr. 1994. "Prison population growth and crime reduction." *Journal of Quantitative Criminology*, 10(2): 109–140.

Mayshar, Joram. 1991. "Taxation with costly administration." *The Scandinavian Journal of Economics*, 75–88.

McArdle, Megan. Mar. 29, 2011. "Did GE Really Pay No U.S. Tax in 2010?" *Atlantic*.

McCollister, Kathryn E, Michael T French, and Hai Fang. 2010. "The cost of crime to society: New crime-specific estimates for policy and program evaluation." *Drug and alcohol dependence*, 108(1): 98–109.

McCormack, John. Nov. 17, 2011. "GE filed 57,000-page tax return, paid no taxes on $14 billion in profits." *The Weekly Standard*.

Mikesell, John, and Maureen A Pirog-Good. 1990. "State lotteries and crime." *American Journal of Economics and Sociology*, 49(1): 7–20.

Mills, Lillian, Merle M. Erickson, and Edward L. Maydew. 1998. "Investments in Tax Planning." *American Taxation Association*, 10: 1.

Ng, Yew-Kwang. 1980. "Optimal Corrective Taxes or Subsidies When Revenue Raising Imposes an Excess Burden." *The American Economic Review*, 744–751.

Pearl, Morris. Sep. 21, 2016. "How Companies Like Apple Dodge Taxes and Their Own Investors." *New York Times.*

Pigou, Arthur C. 1920. "The economics of welfare, 1920." *McMillan&Co., London.*

Pirttilä, Jukka. 2000. "A many-person corlett-hague tax rule with externalities." *Oxford Economic Papers*, 52(3): 595–605.

Pirttilä, Jukka, and Matti Tuomala. 1997. "Income tax, commodity tax and environmental policy." *International Tax and Public Finance*, 4(3): 379–393.

Plott, Charles R. 1966. "Externalities and corrective taxes." *Economica*, 84–87.

Polinsky, A Mitchell, and Steven Shavell. 1992. "Enforcement Costs and the Optimal Magnitude and Probability of Fines." *Journal of Law and Economics*, 133–148.

Polinsky, Mitchell, and Steven Shavell. 1982. "Pigouvian taxation with administrative costs." *Journal of Public Economics*, 19(3): 385–394.

Pollock, Frederick, and Frederic William Maitland. 1899. *The history of English law before the time of Edward I.* Vol. 2, University Press.

Pomp, Richard D. 1993. "The Disclosure of State Corporate Income Tax Data: Turning the Clock Back to the Future." *Capital University Law Review*, 22: 373.

Ramsey, Frank P. 1927. "A Contribution to the Theory of Taxation." *The Economic Journal*, 47–61.

Rice, Eric M. 1991. "The Corporate Tax Gap: Evidence on Tax Compliance by Small Corporations." In *Why People Pay Taxes: Tax Compliance and Enforcement.* , ed. Joel Slemrod, 125–62. University of Michigan Press.

Richman, Daniel C, and William J Stuntz. 2005. "Al Capone's Revenge: An Essay on the Political Economy of Pretextual Prosecution." *Columbia Law Review*, 583–640.

Sandmo, Agnar. 1975. "Optimal taxation in the presence of externalities." *Swedish Journal of Economics*, 86–98.

Sandmo, Agnar. 1978. "Direct versus indirect Pigovian taxation." *European Economic Review*, 7(4): 337–349.

Shaviro, Daniel. 1990. "Beyond public choice and public interest: A study of the legislative process as illustrated by tax legislation in the 1980s." *University of Pennsylvania Law Review*, 139(1): 1–123.

Shepherd, Joanna M. 2002. "Police, Prosecutors, Criminals, and Determinate Sentencing: The Truth about Truth-in-Sentencing Laws." *Journal of Law and Economics*, 45(2): 509–533.

Shepherd, Joanna M. 2004. "Murders of passion, execution delays, and the deterrence of capital punishment." *The Journal of Legal Studies*, 33(2): 283–321.

Sheppard, Lee. 2014. "Should Corporate Tax Returns Be Disclosed?" *Tax Notes*, 142: 1381.

Sherwin, Adam. Nov. 10, 2015. "Crickhowell: Welsh Town Moves 'Offshore' to Avoid Tax on Local Business." *The Independent*.

Slemrod, Joel. 1990. "Optimal Taxation and Optimal Tax Systems." *The Journal of Economic Perspectives*, 4(1): 157–178.

Slemrod, Joel. Mar. 3, 2005. "The Costs of Tax Complexity, Presentation to the President's Advisory Panel on Federal Tax Reform."

Slemrod, Joel, and Shlomo Yitzhaki. 2002. "Tax avoidance, evasion, and administration." *Handbook of Public Economics*, 3: 1423–1470.

Stigler, George J. 1970. "The Optimum Enforcement of Laws." *Journal of Political Economy*, 78(3): pp. 526–536.

Tax Executive Institute. 2006. "TEI Opposes Public Disclosure of Corporate Tax Returns."

Tax Executive Institute. Jul. 25, 2002. "Tax Executives Institute Letter to Treasury, SEC on Corporate Disclosure."

Taxpayer Advocate Service. 2008. "The Complexity of the Tax Code." *Annual Report to Congress*, 1: 13.

The Economist. Sep. 21, 2012. "Corporate Tax Avoidance: The Price Isn't Right."

Wall Street Journal. Jan. 20, 1969. "Treasury Chief Warns of Taxpayer Revolt, Cites $50 Billion a Year in Preferences."

Weber, David P. 2006. "Book-tax differences, analysts' forecast errors, and stock returns." PhD diss. University of Colorado at Boulder.

Yitzhaki, Shlomo. 1979. "A note on optimal taxation and administrative costs." *The American Economic Review*, 475–480.