Optimal Indirect Regulation of Externalities*

Zi Yang Kang†

This version: October 2023

Abstract

This paper studies the regulation of a good that generates different amounts of an externality on consumption, but direct taxation of the externality is infeasible. Under certain conditions, I show that the deadweight loss due to any (possibly nonlinear) tax on the good is equal to the Bregman divergence between the allocation that the tax induces and the first-best allocation. This yields a regression-based method to derive the tax that minimizes deadweight loss in any family of taxes. I use this method to characterize the second-best nonlinear tax and show that quantity restrictions, such as bans and mandates, can be optimal. I quantify the welfare gains of using a nonlinear tax over a linear tax. Finally, I illustrate policy implications by applying my results to the taxation of vehicle miles traveled to regulate automobile externalities.

JEL classification: D47, D62, D63, D82, H23

Keywords: externalities, mechanism design, regression, non-market policy, sufficient statistics

* An earlier version of this paper was circulated under the title "Markets for Goods With Externalities." I am especially indebted to Paul Milgrom, Andy Skrzypacz, and Shosh Vasserman for many illuminating discussions. I also thank Mohammad Akbarpour, Hunt Allcott, Alex Bloedel, Jeremy Bulow, Gabriel Carroll, Piotr Dworczak, Ravi Jagadeesan, Olivier Kooi, Pearl Li, Simon Loertscher, Stephen Morris, Ellen Muir, Afshin Nikzad, Mike Ostrovsky, Ivan Png, Ilya Segal, Takuo Sugaya, Alex Teytelboym, Winnie van Dijk, Bob Wilson, and Frank Yang for helpful comments and suggestions. This paper has additionally benefitted from conference and seminar participants at Stanford and the Markets, Contracts and Organizations Conference 2023. I gratefully acknowledge the SIEPR Program in Regulatory Policy for financial support. An extended abstract of this paper appeared in the *Proceedings of the 24^{th} ACM Conference on Economics and Computation (EC'23)*.

† Center of Mathematical Sciences and Applications, Harvard University; zykang@cmsa.fas.harvard.edu.

1 Introduction

Policymakers often use quantity controls to regulate goods that cause externalities. Psychoactive drugs, for example, are banned in many countries, while immunizations are frequently mandated by schools and employers. Even when outright bans or mandates are not used, quantity restrictions may still apply. Most countries impose compulsory education laws that require children to receive some minimum amount of schooling. One-gun-a-month laws and the Combat Methamphetamine Epidemic Act respectively limit the quantity of firearms and pseudoephedrine any individual can purchase every month.

These policies illustrate how the first-best approach of Pigouvian taxation may not always be practically feasible. In theory, policymakers can ideally induce individuals to internalize the true social impact of their consumption by taxing each individual the marginal external harm that he generates, which restores market efficiency. But in practice, the externality that is generated by recreational drug use, for instance, varies across individuals and is difficult to measure accurately. When policymakers cannot accurately measure (and directly regulate) the externality that each individual causes, they indirectly regulate the externality by regulating the good instead—such as through quantity controls. Unlike a Pigouvian tax, however, indirect policies cannot distinguish between those who generate little social harm and those who generate substantial social harm: all individuals are subject to the same restrictions. Consequently, any indirect policy introduces deadweight loss. How much deadweight loss does each type of indirect policy introduce, and what is the indirect policy that minimizes deadweight loss?

In this paper, I characterize the optimal indirect policy by developing a method to quantify the deadweight loss of any indirect policy. Motivated by the classical Pigouvian analysis, I study a large market in which consumers demand different amounts of a homogeneous good, the consumption of which generates an externality. Unlike the classical Pigouvian analysis, however, I assume that consumers may generate different amounts of externality per unit of good consumed, which the social planner does not observe. As such, the social planner is no longer able to directly regulate the externality. Instead, to indirectly regulate the externality, she chooses a price schedule for the good that might not be linear in the quantity of good consumed. For example, she could impose a linear tax on the good, introduce a quantity ceiling, or even ban the good entirely. Because indirect policies do not distinguish between those who generate high and low amounts of the externality, they distort away from the first-best allocation and result in deadweight loss. The social planner's objective is to minimize deadweight loss.

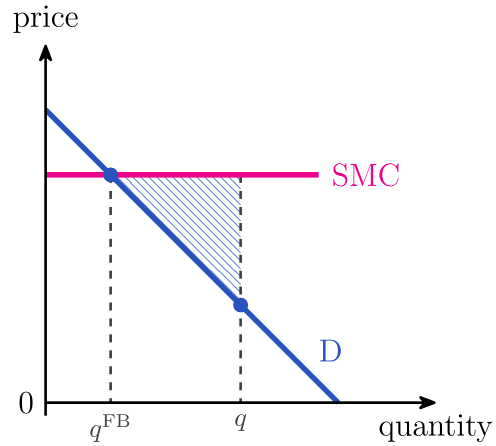

Figure 1: Deadweight loss for an individual consumer with a linear demand curve.

To illustrate my approach, I begin by considering a special case where each individual consumer has a linear demand curve, as illustrated in Figure 1. The first-best allocation, q^{FB}, for a consumer is where his demand curve (D) meets his social marginal cost curve (SMC). At any other allocation q, a deadweight loss arises. The assumption of linear demand simplifies the characterization of deadweight loss and the derivation of the optimal indirect policy.

In this special case, observe that the deadweight loss of any indirect policy is equal to a weighted sum of squared distances between the first-best allocation and the allocation under the indirect policy in question. Indeed, as Harberger (1964) observed, the deadweight loss for an individual consumer is equal to the area of the shaded triangle in Figure 1, which is in turn proportional to the squared distance between q^{FB} and q. The total deadweight loss is thus simply a sum of squared distances between the first-best allocation and the allocation under the indirect policy, suitably weighted to account for how demand and social marginal cost curves differ across consumers.

This observation implies that the problem of finding the deadweight loss-minimizing policy in any set of indirect policies can be equivalently formulated as a *weighted least squares regression* problem. For example, the optimal linear tax can be derived by regressing the externality that each consumer generates per unit of good consumed onto a constant function, which can be computed by applying a standard regression projection formula. More complex nonlinear taxes might require more parameters to be estimated; and the optimal nonlinear tax can be derived by estimating a nonparametric regression. I therefore derive the optimal indirect policy by characterizing the projection formula associated with this nonparametric regression.

Although the assumption of linear demand might appear to play a critical role in this approach, perhaps surprisingly, I show that it does not. In fact, all that is required to generalize this approach beyond linear demand is to employ more general regression loss functions than the weighted sum of squared distances. My first main result shows that the deadweight loss of any indirect policy is more generally equal to a weighted sum of Bregman divergences between the first-best allocation and the allocation under the indirect policy in question. My second main result then shows that the optimal allocation within any set can be derived by estimating a nonparametric regression of the first-best allocation on the set of allocation functions in question, where the regression loss function is a weighted sum of Bregman divergences. This allows me to derive the optimal indirect policy by characterizing the projection operator associated with this nonparametric regression.

Using these results, I show that the optimal indirect policy can require quantity regulations. For instance, suppose that individuals who have the greatest marginal consumption benefit for guns also impose the greatest marginal external harm in expectation. If the external harm that they impose could be directly observed, then only those whose Pigouvian tax exceeds their marginal consumption benefit would be effectively banned from buying guns. But as the external harm that they impose cannot be directly observed, the policymaker cannot distinguish these individuals from others. When the marginal external harm that they impose is sufficiently high, then the optimal indirect policy bans *all* individuals from buying guns. My results allow me to establish a necessary and sufficient condition for the optimal policy to be a ban. When the marginal external harm that they impose is not as high, then the optimal indirect policy might nevertheless set a quantity ceiling on guns to restrict the total external harm that such individuals can impose. In this case, I establish a necessary and sufficient condition for the optimal policy to include a quantity ceiling. Finally, I show that mandates and quantity floors can be similarly optimal when consumption of the good generates a positive, rather than negative, externality.

In addition to characterizing optimal policies, my approach based on regression also uncovers new insights about optimal indirect regulation. First, I show that the social planner's problem can be analogously solved by performing a regression in prices rather than quantities. In particular, the classic formula of the optimal indirect linear tax due to Diamond (1973) can be interpreted as an "estimator" of the social marginal cost in the family of constant marginal taxes. By analogy, the estimators of the social marginal cost in broader families of taxes implement optimal policies in those families, and I derive the tax that implements the optimal indirect policy. Second, I show that applying the law of total variance sheds light on welfare. In particular, I show that the deadweight loss due to the optimal linear tax is equal to the variance of the externality, while the deadweight loss due to the optimal nonlinear tax is equal to the variance of the externality unexplained by consumption. As such, the welfare gain from using the optimal nonlinear tax over the optimal linear tax is equal to the variance of the externality explained by consumption.

The results of this paper provide a simple takeaway for empirical applications: the optimal indirect policy can be derived via a regression approach. To illustrate, I apply my results to the problem of taxing vehicle miles traveled (VMT) to indirectly regulate automobile externalities. This is motivated by recent policy interest in VMT taxation worldwide.[1] On one hand, recent advancements in telematics technology enable policymakers to measure (and hence tax) cumulative VMT. Consequently, policymakers can potentially design taxes that vary nonlinearly in VMT, unlike gas taxes, which are constrained to vary linearly in gas consumption because policymakers cannot measure cumulative gas consumption. On the other hand, the growing number of electric vehicles will eventually render gas taxes ineffective at controlling automobile externalities such as accidents and road damage. By applying my results to the empirical setting of Knittel and Sandler (2018), I examine the use of a VMT tax to indirectly regulate pollutant emissions. Given Knittel and Sandler's estimates of the demand for gasoline and the distribution of pollutant emissions produced per gallon, I show that the optimal indirect policy sets a decreasing marginal tax rate on gasoline, which performs an average of 9.5% better than the optimal linear VMT tax over the sample period.

This paper contributes to the literature on the indirect regulation of externalities by applying tools from mechanism design to derive the optimal nonlinear indirect tax. The seminal work of Diamond (1973) characterized the optimal linear indirect tax, which has been the focus of the subsequent literature—including recent work by Knittel and Sandler (2018), Griffith, O'Connell, and Smith (2019), and Grummon, Lockwood, Taubinsky, and Allcott (2019). However, as argued above, quantity restrictions are used in practice, and technological advancements (in the case of VMT taxation) have made nonlinear indirect taxes feasible.[2] These observations motivate a more general analysis of indirect policies, which this paper provides.[3]

This paper also contributes to a broader literature that aims to understand why non-market policies—such as price and quantity controls—are used in practice. Non-market policies have been shown to justified when policymakers have uncertainty about abatement costs (Weitzman, 1974;

[1] In the U.S., the Bipartisan Infrastructure Bill includes a $125 million pilot study for a VMT tax. A VMT tax is also being considered in the U.K. Countries such as Austria, Belgium, Germany, Russia, and Switzerland have already implemented some form of a VMT tax.

[2] In the literature on law and economics, Fleischer (2015) qualitatively discusses the limitations of using a linear tax on goods to indirectly regulate externalities.

[3] There is an older literature in public finance that studies how direct taxes on externalities interact with nonlinear income taxes (e.g., Sandmo, 1975; Ng, 1980; Bovenberg and van der Ploeg, 1994; Cremer, Gahvari, and Ladoux, 1998; Kopczuk, 2003; Kaplow, 2012). In these contexts, the optimality of nonlinear income taxation arises from asymmetric information about either abatement costs or individual wage rates and tastes, rather than the inability to tax the externality.

Roberts and Spence, 1976; Kaplow and Shavell, 2002), redistributive concerns (Condorelli, 2013; Dworczak ⓡ Kominers ⓡ Akbarpour, 2021; Akbarpour ⓡ Dworczak ⓡ Kominers, 2022; Pai and Strack, 2022), enforcement costs (Glaeser and Shleifer, 2001), and concerns about the equilibrium effects of the policy on the market (Kang, 2023). This paper contributes to this literature by showing that quantity restrictions, such as bans and mandates, can arise from the inability to observe and directly tax the externality.

Finally, the regression approach developed in this paper relates to methods developed for policy evaluation, especially in the sufficient statistics literature. My analysis is closest to Jacobsen, Knittel, Sallee, and van Benthem (2020), who build on the pioneering work of Harberger (1964) to show how regression statistics can be used to measure deadweight loss in indirect policies. While Jacobsen et al. focus on linear taxation in the case of a linear demand system, I demonstrate how a regression approach nonetheless applies to general indirect policies in the case of general demand systems. By contrast to sufficient statistics that apply to small changes in policies, this generalization allows the regression approach to apply even for large changes in policies.

The remainder of this paper is organized as follows. In Section 2, I develop a model of indirect regulation. Section 3 then illustrates the regression approach and builds intuition for the general analysis by considering the special case where consumers have linear demand. The general analysis and theoretical results are provided in Section 4; and the economic implications of my results on implementation, welfare, and the optimality of quantity restrictions are discussed in Section 5. I apply the regression approach to the empirical problem of VMT taxation in Section 6 and discuss extensions in Section 7. Finally, Section 8 concludes.

2 Model

In this section, I develop a model in which consumers generate different amounts of an externality from consuming a good. I then analyze the first-best benchmark and formulate the social planner's problem of indirectly regulating the externality by regulating the good.

2.1 Setup

There is a unit mass of risk-neutral consumers in a market for a homogeneous good. Consumption of the good generates an externality. Consumers differ in their consumption preferences and the amounts of externality that they generate from consumption. The good is supplied competitively at a constant marginal cost c, and each consumer demands up to $A > 0$ units of the good.

Heterogeneity in consumption preferences is captured by a *consumption type* $\theta \in [\underline{\theta}, \overline{\theta}] \subset \mathbb{R}_+$ about which the consumer is privately informed. This determines the utility $\theta v(q)$ that he derives from consuming a quantity q of the good, where $v : [0, A] \to \mathbb{R}_+$ is bounded, twice continuously differentiable, and satisfies standard assumptions of strict monotonicity and decreasing marginal utility: $v'(q) > 0$ and $v''(q) < 0$ for any $q \in [0, A]$. Each consumer's individual demand curve $D(\cdot\,; \theta) : \mathbb{R} \to [0, A]$ can thus be written as

$$D(p; \theta) = (v')^{-1} \left(\frac{p}{\theta} \right).$$

Heterogeneity in externality generated is captured by an *externality type* $\xi \in [\underline{\xi}, \overline{\xi}] \subset \mathbb{R}$. This is the marginal externality (measured in dollar terms) that the consumer generates per unit of good consumed (see Section 7.1 for a discussion of the case when the generated externality varies nonlinearly with the amount of good consumed). I assume that consumers are also privately informed about ξ; all results are identical if consumers do not know their own externality types (see Lemma 1 below). The joint distribution of (θ, ξ) is denoted by G, which is absolutely continuous and has a positive density function g on $[\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$.

Each consumer's utility depends both on his own consumption of the good and the aggregate externality generated by all consumers. Following standard textbook analyses, I focus on the case where utility is additively separable in the aggregate externality (again, discussion of the general case is deferred to Section 7.1). Thus, if a consumer with type (θ, ξ) consumes a quantity $q(\theta, \xi)$ of the good and makes a payment of $t(\theta, \xi)$, then he realizes a utility of

$$\theta v(q(\theta, \xi)) - t(\theta, \xi) - E, \quad \text{where } E = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \xi q(\theta', \xi') \, \mathrm{d}G(\theta', \xi').$$

For simplicity, I assume throughout the paper that $\underline{\xi} \geq 0$ unless otherwise stated. This allows me to interpret the externality as a negative externality. All results extend easily to positive externalities by allowing $\underline{\xi} < 0$ (and in fact different consumers can generate externalities of different signs).

2.2 First-best benchmark

Before introducing the social planner's problem, I briefly analyze the first-best benchmark. To this end, suppose that the externality that each consumer generates from consumption can be measured and directly regulated. Following the analysis of Pigou (1920), the first-best outcome is attained by setting a personalized Pigouvian tax equal to ξ per unit of good consumed. Consequently, I define the *first-best allocation function* $q^{\mathrm{FB}} : [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}] \to [0, A]$ by

$$q^{\mathrm{FB}}(\theta, \xi) = D(c + \xi; \theta).$$

However, the first-best benchmark may not be attainable in practice. As argued in Section 1, policymakers often cannot measure and directly regulate the externality due to technological challenges. When direct regulation of the externality is infeasible, the externality can at best be indirectly regulated by regulating the good instead.

2.3 Social planner's problem

I now formulate the social planner's problem of indirectly regulating the externality by regulating the good. To regulate the good, the social planner chooses a mechanism (q, t), which consists of an allocation function $q : [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}] \to [0, A]$ and a payment function $t : [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}] \to \mathbb{R}$. These respectively determine the quantity of the good that each consumer consumes and the payment that he makes to the social planner.

I proceed by first characterizing the set of implementable allocation functions and then stating the social planner's objective.

Implementable allocation functions. By the revelation principle, it is without any loss of generality for the social planner to consider only mechanisms under which consumers report their consumption and externality types truthfully. Formally, the social planner restricts attention to mechanisms (q, t) that satisfies the following incentive compatibility constraint:

$$\text{for any } (\theta, \xi), \qquad (\theta, \xi) \in \arg\max_{(\hat{\theta}, \hat{\xi})} \left[\theta v(q(\hat{\theta}, \hat{\xi})) - t(\hat{\theta}, \hat{\xi}) - \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \xi' q(\theta', \xi') \, \mathrm{d}G(\theta', \xi') \right]. \tag{IC}$$

Notice that, because of the large market assumption, individual reports $(\hat{\theta}, \hat{\xi})$ do not affect the aggregate externality in the market; hence (as in the standard Pigouvian analysis) each consumer fails to internalize how much his own consumption affects others.

Next, I characterize the set of allocation functions that the social planner considers. Following the mechanism design literature, I call an allocation function q *implementable* if there exists a payment function $t : [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}] \to \mathbb{R}$ such that the mechanism (q, t) satisfies (IC). The following lemma characterizes the set of implementable allocation functions:

Lemma 1. *Define*

$$\mathcal{Q} := \left\{ q : [\underline{\theta}, \overline{\theta}] \to [0, A] \text{ is non-decreasing} \right\}.$$

Then an allocation function q is implementable only if there exists $\hat{q} \in \mathcal{Q}$ such that $q(\theta, \xi) = \hat{q}(\theta)$ for almost every $(\theta, \xi) \in [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$.

Lemma 1 shows that the formulation—and thus solution—of the social planner's problem does not depend on whether consumers observe their externality types. If consumers do not observe ξ, then the social planner can clearly do no better than to elicit only information about θ. However, even if consumers observe ξ, Lemma 1 shows that the social planner can also do no better than to elicit only information about θ: any implementable allocation function must be independent of ξ for almost every consumer. Henceforth, I refer to $\mathcal{Q}$ defined in Lemma 1 as the set of implementable allocation functions; I also write q and t in any incentive-compatible mechanism as functions of only θ.

While Lemma 1 is formally proven in Appendix A, its underlying intuition is straightforward: consumption choices are determined entirely by consumption types θ, rather than externality types ξ. This can be seen from (IC): although consumers can report ξ, their reports do not directly affect their payoffs. Consequently, no incentive-compatible mechanism can truthfully elicit information about ξ; instead, the social planner must form beliefs about ξ based on only consumers' revealed consumption preferences.[4] In addition, implementable allocation functions must be non-decreasing in θ due to an adaptation of Myerson's (1981) lemma to the present setting.

Social planner's objective. The social planner is utilitarian and aims to maximize total social surplus. Under any incentive-compatible mechanism (q, t), total social surplus can be written as

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \left[\theta v(q(\theta)) - (c + \xi)\, q(\theta)\right] \, \mathrm{d}G(\theta, \xi).$$

Notice that total social surplus does not directly depend on the payment function t as cash transfers between consumers and the social planner are welfare-neutral. Equivalently, the payments that the social planner receives from indirectly regulating the externality are "recycled" and distributed evenly among all consumers. I discuss revenue and distributional concerns in Section 7.2.

The social planner's problem can be equivalently formulated as one of minimizing deadweight loss. Under any incentive-compatible mechanism (q, t), deadweight loss is equal to the difference

[4] A similar intuition underlies analogous characterization results in the mechanism design literature, such as those by Jehiel and Moldovanu (2001), Che, Dessein, and Kartik (2013), and Dworczak Ⓡ al. (2021).

between the total social surplus under the first-best allocation function q^{FB} and the total social surplus under the allocation function q:

$$\text{DWL} = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \left\{\theta \left[v(q^{\text{FB}}(\theta, \xi)) - v(q(\theta))\right] - (c + \xi) \left[q^{\text{FB}}(\theta, \xi) - q(\theta)\right]\right\} \, \mathrm{d}G(\theta, \xi).$$

To solve for the optimal mechanism, the social planner thus chooses an implementable allocation function $q \in \mathcal{Q}$ to minimize the above expression. Henceforth, I use this formulation of the social planner's problem instead so as to facilitate stating my results, which characterize and minimize deadweight loss to derive the optimal mechanism via a regression approach.

3 Illustration of approach

I now illustrate how regression can be used to measure the deadweight loss of any mechanism and derive the optimal mechanism. To this end, I consider the following special case in this section and defer the analysis of the general case to Section 4:

Assumption 1. *Consumers have quadratic utility given by $v(q) = Aq - q^2/2$ for $q \in [0, A]$, where all consumers derive sufficiently high marginal utility for the initial unit of the good: $\underline{\theta}A > c + \overline{\xi}$.*

Under Assumption 1, all consumers have linear demand curves and consume a positive amount of the good under the first-best benchmark. In particular, consumers' individual demand curves are given by

$$D(p; \theta) = A - \frac{p}{\theta}.$$

Because of the condition on consumers' marginal utility for the initial unit of the good, this implies that the first-best allocation function is positive:

$$q^{\text{FB}}(\theta, \xi) = D(c + \xi; \theta) = A - \frac{c + \xi}{\theta} > 0 \qquad \text{for any } (\theta, \xi) \in [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}].$$

The purpose of Assumption 1 is twofold. On one hand, it simplifies the problem and builds intuition for the general analysis in Section 4. On the other hand, it also serves as an approximation in the spirit of Harberger (1964) and connects my approach to the sufficient statistics literature.

3.1 Deadweight loss

I begin by showing how deadweight loss is equal to a weighted sum of the squared distances between the first-best allocation function and the allocation function of a given mechanism.

Proposition 1. *Under Assumption 1, the deadweight loss due to any mechanism (q,t) is equal to half of the average squared distance between the allocation function q and the first-best allocation function q^{FB}, weighted by consumption types θ:*

$$\mathrm{DWL} = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \frac{\theta}{2} \left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right]^2 \, \mathrm{d}G(\theta,\xi).$$

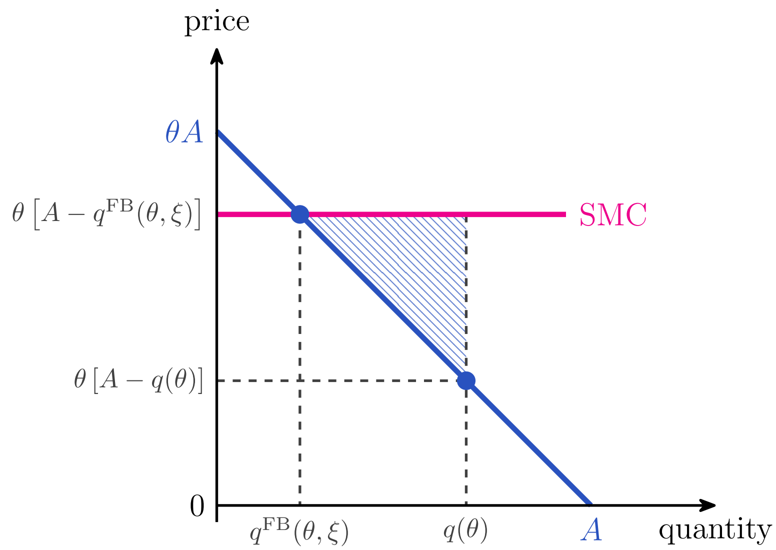

Figure 2: Deadweight loss due to a mechanism (q,t) for an individual consumer.

The expression for deadweight loss in Proposition 1 follows straightforwardly from the familiar analysis of Harberger triangles. Indeed, for a consumer with consumption type θ and externality type ξ, the deadweight loss of consuming $q(\theta)$ units of the good is equal to the area of the shaded triangle in Figure 2:

$$\mathrm{DWL}(\theta,\xi) = \frac{1}{2} \times \left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right] \times \theta\left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right] = \frac{\theta}{2}\left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right]^2.$$

By summing this over all consumers, the expression for deadweight loss in Proposition 1 can thus be obtained.

Despite its simplicity, Proposition 1 uncovers the following key relationship between deadweight loss and regression:

Corollary 1. *Under Assumption 1, the allocation function that minimizes deadweight loss within any set of implementable allocation functions is equal to the weighted least squares projection (with weights equal to θ) of the first-best allocation function q^{FB} onto that set.*

Corollary 1 shows that minimizing deadweight loss within any set of implementable allocation functions is equivalent to performing a weighted least squares regression. To see why, consider a set $\mathcal{S} \subset \mathcal{Q}$ of allocation functions and a weighted least squares regression of the first-best allocation function q^{FB} onto allocation functions in $\mathcal{S}$, with weights equal to consumption types θ:

$$q^{\mathrm{FB}}(\theta, \xi) = q(\theta) + \varepsilon(\theta, \xi), \qquad q \in \mathcal{S}.$$

The sum of squared residuals of this weighted least squares regression is

$$\begin{aligned} \mathrm{SSR} &= \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \left[\varepsilon(\theta, \xi)\right]^2 \, \mathrm{d}G(\theta, \xi) \\ &= \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \left[q^{\mathrm{FB}}(\theta, \xi) - q(\theta)\right]^2 \, \mathrm{d}G(\theta, \xi) = 2 \times \mathrm{DWL}\,. \end{aligned}$$

Consequently, the solution to the above weighted least squares regression simultaneously minimizes both the sum of squared residuals and deadweight loss over all allocation functions in $\mathcal{S}$.

In analogy with linear regression problems, this weighted least squares regression problem can be solved by applying an appropriate projection operator to the first-best allocation function q^{FB}. In a linear regression, the best-fit parameters are obtained by projecting the data onto the linear subspace spanned by the covariates; thus the solution to a linear regression problem can be found by applying a projection matrix to the data. Similarly, the optimal allocation function within any set of implementable allocation functions can be found by projecting the first-best allocation function q^{FB} onto that set.

Corollary 1 can therefore be used to find the optimal allocation within any set of implementable allocation functions. As a first illustration, consider the problem of finding the optimal linear tax, as in Diamond (1973). The optimal linear tax can then be found by minimizing the deadweight loss over the set of all allocation functions that can be implemented by a constant marginal tax τ,

$$\mathcal{S} = \left\{ q : [\underline{\theta}, \overline{\theta}] \to [0, A] \text{ such that } q(\theta) = A - \frac{c + \tau}{\theta} \text{ for some } \tau \in \mathbb{R} \right\}.$$

By Corollary 1, the optimal constant marginal tax can be determined via the weighted least squares projection of q^{FB} onto $\mathcal{S}$:

$$\underbrace{A - \frac{c+\xi}{\theta}}_{=q^{\mathrm{FB}}(\theta,\xi)} = A - \frac{c+\tau}{\theta} + \varepsilon(\theta,\xi), \qquad \tau \in \mathbb{R}.$$

Equivalently, the optimal constant marginal tax can be derived by regressing the amounts of externality that consumers generate onto the set of constant functions:

$$\xi = \tau - \theta \cdot \varepsilon(\theta,\xi), \qquad \tau \in \mathbb{R}.$$

Using the textbook formula for the regression coefficient for such a linear regression (e.g., Exercise 5 in Section 3.7 of James, Witten, Hastie, and Tibshirani, 2021), the optimal constant marginal tax is

$$\tau^* = \frac{\mathbf{E}\left[\xi/\theta\right]}{\mathbf{E}\left[1/\theta\right]}.$$

This recovers the familiar result of Diamond (1973) for the case of linear demand: the optimal constant marginal tax on the good is equal to the weighted average of the marginal harm imposed by each consumer, where the weights are given by the slopes of consumers' individual demand curves, $1/\theta$.

3.2 Second-best allocation

I now show how Corollary 1 can be used to derive the optimal implementable allocation function, which I refer to as the second-best allocation function, q^{SB}.

I begin by using the result of Proposition 1 to rewrite the social planner's problem of minimizing deadweight loss as a regression problem. To this end, observe that q^{SB} solves

$$\min_{q\in\mathcal{Q}} \int_{\underline{\theta}}^{\bar{\theta}} \int_{\underline{\xi}}^{\bar{\xi}} \frac{\theta}{2}\left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right]^2 \, \mathrm{d}G(\theta,\xi).$$

By the discussion following Corollary 1, this can be viewed as a weighted least squares regression problem: the social planner chooses an implementable allocation function $q \in \mathcal{Q}$ that minimizes the weighted sum of squared distances to the first-best allocation function q^{FB}. Therefore, to solve this regression problem, it suffices to characterize the associated projection operator.

I now characterize the projection operator associated with this regression. To this end, let the concave closure of a function ϕ be defined as the pointwise smallest concave function that lies above ϕ, and denote the concave closure of ϕ by $\operatorname{co}\phi$. The following lemma provides the desired characterization:

Lemma 2. *Let $q_0 : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}$ and $w : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}_{++}$ be two functions. Denote $W(\theta) := \int_{\underline{\theta}}^{\theta} w(z)\ \mathrm{d}z$ and suppose that $W(\overline{\theta}) = 1$. Then there exists a unique solution (up to equality almost everywhere) to the problem*

$$\min_{q \in \mathcal{Q}} \int_{\underline{\theta}}^{\overline{\theta}} [q_0(\theta) - q(\theta)]^2\, w(\theta)\ \mathrm{d}\theta.$$

Moreover, the unique solution is given by

$$q^*(\theta) = \begin{cases} 0 & \text{if } \overline{q}_0(\theta) \leq 0, \\ A & \text{if } \overline{q}_0(\theta) \geq A, \\ \overline{q}_0(\theta) & \text{otherwise,} \end{cases} \quad \text{where } \overline{q}_0(\theta) := \frac{\mathrm{d}}{\mathrm{d}s}\left(\operatorname{co}\int_{W^{-1}(s)}^{\overline{\theta}} q_0(z)w(z)\ \mathrm{d}z\right)\Bigg|_{s=W(\theta)}.$$

The proof of Lemma 2 is deferred to Appendix A due to its technical nature. Instead, I explain the content of Lemma 2 below in three parts.

First, Lemma 2 shows that there is a *unique* implementable allocation function that is closest to the any given function q_0 (in the sense of weighted sum of squared distances), which allows the projection operator to be well-defined. On one hand, existence follows straightforwardly from the fact that the objective functional is continuous in q and that the set $\mathcal{Q}$ is closed. On the other hand, uniqueness follows from the fact that the objective functional is strictly convex in q: if there were two allocation functions in $\mathcal{Q}$ that minimized the objective functional, then their arithmetic average would attain an even lower value, leading to a contradiction. These observations imply that there is a unique allocation function $q^* \in \mathcal{Q}$ that is closest to the given function q_0; thus q^* is the projection of q_0 onto $\mathcal{Q}$.

Second, Lemma 2 shows that the projection of a given function q_0 onto $\mathcal{Q}$ requires *ironing* the function if it is not already non-decreasing. Indeed, if q_0 is already non-decreasing, then

$$\operatorname{co}\int_{W^{-1}(s)}^{\overline{\theta}} q_0(z)w(z)\ \mathrm{d}z = \int_{W^{-1}(s)}^{\overline{\theta}} q_0(z)w(z)\ \mathrm{d}z \implies \overline{q}_0(\theta) = q_0(\theta).$$

However, if q_0 is not already non-decreasing, then it must be ironed by leveling non-monotone regions via a weighted equal-areas condition, as illustrated in Figure 3. More precisely, in each ironing interval, the ironed function $\overline{q}_0$ must be constant such that the integrals of $q_0(\theta)$ and $\overline{q}_0(\theta)$,

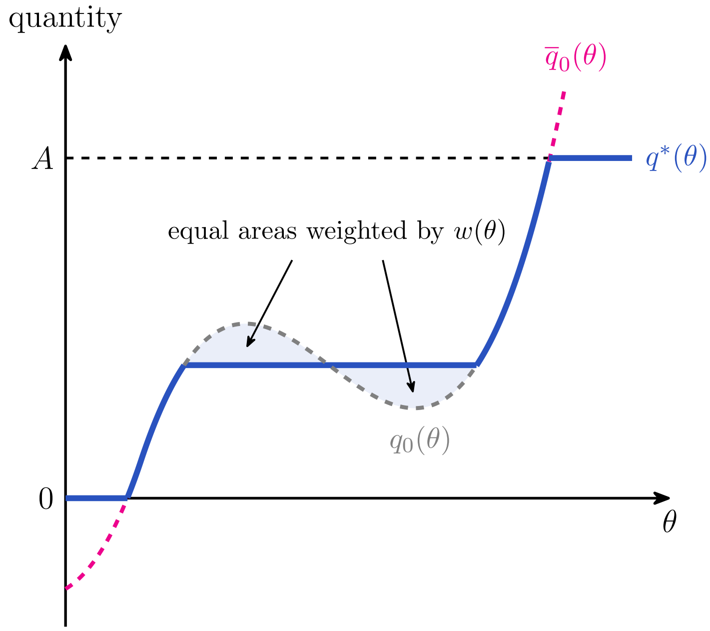

Figure 3: Projection of a given function q_0 characterized by Lemma 2.

each weighted by $w(\theta)$, are equal. Outside of these ironing intervals (where q_0 is non-decreasing), $\overline{q}_0$ and q_0 must coincide.

Third, Lemma 2 shows that the resulting ironed function $\overline{q}_0$ might still require *truncation* if it results in infeasible allocations of the good. As illustrated in Figure 3, the image of q_0, and hence the image of $\overline{q}_0$, can lie outside of the feasible interval $[0, A]$. To obtain an implementable allocation function in $\mathcal{Q}$, the ironed function $\overline{q}_0$ must be truncated below 0 and above A, so that all consumption types consume a feasible allocation of the good.

Finally, I solve the social planner's regression problem by applying the projection operator characterized in Lemma 2. To this end, let the functions q_0 and w in the statement of Lemma 2 be defined by

$$q_0(\theta) = \int_{\underline{\xi}}^{\overline{\xi}} q^{\mathrm{FB}}(\theta, \xi) g(\theta, \xi) \, \mathrm{d}\xi \quad \text{and} \quad w(\theta) = \frac{\theta}{\mathbf{E}[\theta]} \int_{\underline{\xi}}^{\overline{\xi}} g(\theta, \xi) \, \mathrm{d}\xi.$$

Moreover, Assumption 1 implies that the first-best allocation function is feasible for all consumers since the image of q^{FB} is interior: $\operatorname{im} q^{\mathrm{FB}} \in (0, A)$. This means that only ironing might be required under Assumption 1; truncation is not. The following proposition summarizes this discussion:

Proposition 2. *Let W be the absolutely continuous distribution with positive density w on the interval $[\underline{\theta}, \overline{\theta}]$ defined by*

$$w(\theta) = \frac{\theta}{\mathbf{E}[\theta]} \int_{\underline{\xi}}^{\overline{\xi}} g(\theta, \xi) \, \mathrm{d}\xi \quad \textit{and} \quad W(\theta) = \frac{1}{\mathbf{E}[\theta]} \int_{\underline{\theta}}^{\theta} \int_{\underline{\xi}}^{\overline{\xi}} z g(z, \xi) \, \mathrm{d}\xi \, \mathrm{d}z.$$

Under Assumption 1, the second-best allocation function q^{SB} is unique and equal to

$$q^{\mathrm{SB}}(\theta) = \frac{\mathrm{d}}{\mathrm{d}s} \left(\mathrm{co} \int_{s}^{1} \mathbf{E} \left[q^{\mathrm{FB}}(\hat{\theta}, \xi) \, | \, \hat{\theta} = W^{-1}(z) \right] \, \mathrm{d}z \right) \bigg|_{s = W(\theta)}.$$

4 Theoretical results

Although the insights derived in the previous section rely on the assumption of linear demand, I now show in this section how they extend to the general model.

4.1 Deadweight loss

I begin by demonstrating how the connection between deadweight loss and regression nonetheless extends to the general model. To this end, I show that deadweight loss is equal to a weighted sum of Bregman divergences associated with $-v$ between the first-best allocation function and the allocation function of a given mechanism, rather than a weighted sum of squared distances under the assumption of linear demand (cf. Assumption 1). Recall that the generalized Bregman divergence Δ_Φ associated with a strictly convex function $\Phi : \Omega \to \mathbb{R}$ defined on a closed convex domain $\Omega \subset \mathbb{R}$ with a selection of subgradients $\phi \in \partial \Phi$ is given by

$$\Delta_\Phi(q_1, q_2) = \Phi(q_1) - \Phi(q_2) - (q_1 - q_2) \, \phi(q_2) \qquad \text{for } q_1, q_2 \in \Omega.$$

Below, I consider the following Bregman divergence associated with the strictly convex function $-v : [0, A] \to \mathbb{R}$ for a consumer of type (θ, ξ):

$$\Delta_{-v}^{(\theta,\xi)}(q_1, q_2) := v(q_2) - v(q_1) - (q_1 - q_2) \, \phi^{(\theta,\xi)}(q_2), \quad \text{where } \phi^{(\theta,\xi)}(q) = \begin{cases} -v'(q) & \text{for } q \in (0, A), \\ -\dfrac{c + \xi}{\theta} & \text{for } q = 0 \text{ or } A. \end{cases}$$

Henceforth, I refer to $\Delta_{-v}^{(\theta,\xi)}$ as the Bregman divergence (without qualification) since no confusion should arise. The above construction of the Bregman divergence ensures that the subgradient function $\phi^{(\theta,\xi)}$ always satisfies

$$-\theta\phi^{(\theta,\xi)}(q^{\mathrm{FB}}(\theta,\xi)) = c+\xi. \tag{1}$$

Indeed, when $q^{\mathrm{FB}}(\theta,\xi) \in (0,A)$ is interior, then $\phi^{(\theta,\xi)}(q^{\mathrm{FB}}(\theta,\xi)) = -v'(q^{\mathrm{FB}}(\theta,\xi))$ since $\phi^{(\theta,\xi)}$ must coincide with the unique subgradient $-v'$ on the interval $(0,A)$; thus equation (1) reduces to the first-order condition

$$\theta v'(q^{\mathrm{FB}}(\theta,\xi)) = c+\xi.$$

However, when $q^{\mathrm{FB}}(\theta,\xi) = 0$ or A, then $-v$ is no longer differentiable and has infinitely many subgradients. The above selection of $\phi^{(\theta,\xi)}$ at $q=0$ and A ensure that equation (1) is nonetheless satisfied.

Theorem 1. *The deadweight loss due to any mechanism (q,t) is equal to the average Bregman divergence between the allocation function q and the first-best allocation function q^{FB}, weighted by consumption types θ:*

$$\mathrm{DWL} = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)} \left(q(\theta), q^{\mathrm{FB}}(\theta,\xi)\right) \, \mathrm{d}G(\theta,\xi).$$

Proof. For a given consumer of type (θ,ξ), observe that the deadweight loss due to any mechanism (q,t) for that consumer can be written as

$$\begin{aligned}
\mathrm{DWL}(\theta,\xi) &= \theta\left[v(q^{\mathrm{FB}}(\theta,\xi)) - v(q(\theta))\right] - (c+\xi)\left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right] \\
&= \theta\left[v(q^{\mathrm{FB}}(\theta,\xi)) - v(q(\theta))\right] - \theta\phi^{(\theta,\xi)}(q^{\mathrm{FB}}(\theta,\xi))\left[q^{\mathrm{FB}}(\theta,\xi) - q(\theta)\right] \\
&= \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{FB}}(\theta,\xi)\right).
\end{aligned}$$

The second equality follows from equation (1) and the third equality follows from the definition of the Bregman divergence $\Delta_{-v}^{(\theta,\xi)}$. The desired expression for deadweight loss follows by summing over all consumers. □

Theorem 1 extends Proposition 1 by generalizing the measure of distance between the allocation function and the first-best allocation function. To see why, observe that under Assumption 1, the Bregman divergence between $q(\theta)$ and $q^{\mathrm{FB}}(\theta,\xi)$ is equal to half of the squared distance between the two functions:

$$\begin{aligned}\Delta_{-v}^{(\theta,\xi)}\left(q(\theta),q^{\mathrm{FB}}(\theta,\xi)\right) &= -Aq(\theta)+\frac{1}{2}\left[q(\theta)\right]^2+Aq^{\mathrm{FB}}(\theta,\xi)-\frac{1}{2}\left[q^{\mathrm{FB}}(\theta,\xi)\right]^2 \\ &\quad+\left[q(\theta)-q^{\mathrm{FB}}(\theta,\xi)\right]\left[A-q^{\mathrm{FB}}(\theta,\xi)\right] \\ &= \frac{1}{2}\left[q^{\mathrm{FB}}(\theta,\xi)-q(\theta)\right]^2.\end{aligned}$$

This generalization implies that, as in the case of linear demand, the optimal allocation function in any set of implementable allocation functions can be found via a regression approach—simply by generalizing the regression loss function. Indeed, for any set $\mathcal{S}\subset\mathcal{Q}$ of allocation functions, we can define the *weighted Bregman projection* of q^{FB} onto $\mathcal{S}$ by

$$\operatorname*{arg\,min}_{q\in\mathcal{S}}\int_{\underline{\theta}}^{\overline{\theta}}\int_{\underline{\xi}}^{\overline{\xi}}\theta\cdot\Delta_{-v}^{(\theta,\xi)}\left(q(\theta),q^{\mathrm{FB}}(\theta,\xi)\right)\ \mathrm{d}G(\theta,\xi).$$

In turn, this observation generalizes Corollary 1 by showing how optimal allocation functions in the general model can nonetheless be similarly characterized as solutions to regression problems.

4.2 Second-best allocation

I now use Theorem 1 to derive the second-best allocation function.

Theorem 2. *Let W be the absolutely continuous distribution with positive density w on the interval $[\underline{\theta},\overline{\theta}]$ defined by*

$$w(\theta)=\frac{\theta}{\mathbf{E}[\theta]}\int_{\underline{\xi}}^{\overline{\xi}}g(\theta,\xi)\ \mathrm{d}\xi\quad\text{and}\quad W(\theta)=\frac{1}{\mathbf{E}[\theta]}\int_{\underline{\theta}}^{\theta}\int_{\underline{\xi}}^{\overline{\xi}}zg(z,\xi)\ \mathrm{d}\xi\ \mathrm{d}z.$$

Moreover, define

$$\overline{q}^{\mathrm{FB}}(\theta):=(v')^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_0^s\frac{c+\mathbf{E}\left[\xi\mid\theta=W^{-1}(z)\right]}{W^{-1}(z)}\ \mathrm{d}z\right)\Bigg|_{s=W(\theta)}\right).$$

Then the second-best allocation function q^{SB} *is unique and equal to*

$$q^{\mathrm{SB}}(\theta) = \begin{cases} 0 & \textit{if } \overline{q}^{\mathrm{FB}}(\theta) \le 0, \\ A & \textit{if } \overline{q}^{\mathrm{FB}}(\theta) \ge A, \\ \overline{q}^{\mathrm{FB}}(\theta) & \textit{otherwise.} \end{cases}$$

Theorem 2 extends Proposition 2 by showing that, even in the general model, the second-best allocation function is unique and can be obtained through a combination of ironing and truncation (cf. Lemma 2). Indeed, the function $\overline{q}^{\mathrm{FB}}$ in the statement of the theorem can be interpreted as an "ironed first-best allocation function" averaged over externality types. To see why, observe that under Assumption 1, because $v'(q) = A - q$, $\overline{q}^{\mathrm{FB}}$ can be written as

$$\begin{aligned}
\overline{q}^{\mathrm{FB}}(\theta) &= A - \frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_0^s \frac{c + \mathbf{E}\left[\xi \,|\, \theta = W^{-1}(z)\right]}{W^{-1}(z)}\ \mathrm{d}z\right)\bigg|_{s=W(\theta)} \\
&= \frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_s^1 \left[A - \frac{c + \mathbf{E}\left[\xi \,|\, \theta = W^{-1}(z)\right]}{W^{-1}(z)}\right]\ \mathrm{d}z\right)\bigg|_{s=W(\theta)} \\
&= \frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_s^1 \mathbf{E}\left[q^{\mathrm{FB}}(\hat{\theta}, \xi) \,|\, \hat{\theta} = W^{-1}(z)\right]\ \mathrm{d}z\right)\bigg|_{s=W(\theta)}.
\end{aligned}$$

A similar computation for general v justifies this interpretation. Thus the second-best allocation function must be equal to the ironed first-best allocation function $\overline{q}^{\mathrm{FB}}$, truncated below 0 and above A to ensure feasibility.

The key to proving Theorem 2 lies in the following lemma, which builds on Lemma 2 in order to characterize q^{SB} as the projection of q^{FB} onto the set of implementable allocation functions $\mathcal{Q}$. As its proof is technical, it is deferred to Appendix A.

Lemma 3. *Let* q^{SB} *denote the second-best allocation function as defined in Theorem 2. For any implementable allocation function* $q \in \mathcal{Q}$*, the following inequality holds:*

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{FB}}(\theta,\xi)\right)\ \mathrm{d}G(\theta,\xi) \\
&\ge \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{SB}}(\theta)\right)\ \mathrm{d}G(\theta,\xi) + \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q^{\mathrm{SB}}(\theta), q^{\mathrm{FB}}(\theta,\xi)\right)\ \mathrm{d}G(\theta,\xi).
\end{aligned}$$

To complete the proof of Theorem 2, notice that the Bregman divergence between any two quantities $q_1, q_2 \in [0, A]$ is non-negative due to the concavity of v: since $\phi^{(\theta,\xi)}(q_2) = -v'(q_2)$ for almost every $q_2 \in [0, A]$,

$$\begin{aligned}\Delta_{-v}^{(\theta,\xi)}(q_1, q_2) &= v(q_2) - v(q_1) - (q_1 - q_2)\,\phi^{(\theta,\xi)}(q_2) \\ &= \int_{q_1}^{q_2} \left[v'(s) - v'(q_2)\right]\ \mathrm{d}s = -\int_{q_1}^{q_2}\int_{s}^{q_2} v''(z)\ \mathrm{d}z\ \mathrm{d}s \geq 0.\end{aligned}$$

Thus Lemma 3 implies that the minimum deadweight loss must be no smaller than the deadweight loss attained by q^{SB}:

$$\min_{q\in\mathcal{Q}} \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{FB}}(\theta,\xi)\right)\ \mathrm{d}G(\theta,\xi) \geq \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \Delta_{-v}^{(\theta,\xi)}\left(q^{\mathrm{SB}}(\theta), q^{\mathrm{FB}}(\theta,\xi)\right)\ \mathrm{d}G(\theta,\xi).$$

However, it is easily verified that $q^{\mathrm{SB}} \in \mathcal{Q}$ is an implementable allocation function. Consequently, q^{SB} minimizes the deadweight loss among all implementable allocation functions.

5 Economic implications

In this section, I use the results derived in Section 4 to analyze the economic implications of indirect regulation of externalities. First, I demonstrate how the regression approach sheds light on how optimal allocation in any set can be implemented. Second, I study the welfare implications of the regression approach. Finally, I analyze when quantity restrictions, including bans and mandates, are optimal.

5.1 Implementation of optimal allocation

To implement the optimal allocation, the first result of this section shows that the corresponding marginal price faced by each consumer can be similarly derived by performing a suitable regression.

For any incentive-compatible mechanism (q, t), define the marginal price $p(\theta)$ as the marginal price that a consumer with consumption type θ faces. On one hand, the marginal price function can be derived by first computing the payment function by applying the envelope theorem to the optimal allocation function. On the other hand, the regression approach that I explain below is more direct and insightful. In particular, it leads to a reinterpretation and generalization of Diamond's (1973) formula for the optimal linear tax.

To derive the regression for the optimal marginal price, I consider the Bregman divergence associated with the consumer surplus function. The consumer surplus function $\mathrm{CS}(\cdot;\theta) : [0, \theta A] \to$

$\mathbb{R}$ for each consumer is strictly convex and defined by

$$\mathrm{CS}(p;\theta) = \int_p^{\theta A} D(s;\theta)\ \mathrm{d}s = \int_p^{\theta A} (v')^{-1}\left(\frac{s}{\theta}\right)\ \mathrm{d}s.$$

The associated Bregman divergence for each consumer is therefore defined by

$$\Delta^{\theta}_{\mathrm{CS}}(p_1, p_2) := \mathrm{CS}(p_2;\theta) - \mathrm{CS}(p_1;\theta) + (p_1 - p_2)\,(v')^{-1}\left(\frac{p_2}{\theta}\right).$$

Proposition 3. *Given an incentive-compatible mechanism (q,t), let the marginal price function be p. Then the deadweight loss due to (q,t) is equal to the average Bregman divergence associated with $\mathrm{CS}(\cdot;\theta)$ between the marginal social cost $c+\xi$ and the marginal price function p:*

$$\mathrm{DWL} = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, p(\theta)\right)\ \mathrm{d}G(\theta,\xi).$$

Consequently, given any set $\mathcal{S}$ of marginal price functions, the marginal price function in $\mathcal{S}$ that minimizes deadweight loss can be found by projecting the marginal social cost $c+\xi$ onto $\mathcal{S}$.

Proposition 3 shows that the regression approach implied by Theorem 1 applies not only to quantities, but also to prices as well. In particular, the marginal price function that minimizes deadweight loss in any given set can be found by performing a regression that projects the marginal social cost onto that set, just as the optimal allocation function in any given set can be found by performing a regression that projects the first-best allocation function onto that set. Although the loss functions for the two regressions are naturally different, Proposition 3 shows that the loss function for the price regression remains an average of Bregman divergences.

Proposition 3 leads to a reinterpretation and generalization of Diamond's (1973) well-known formula for the optimal constant marginal tax. Consider the marginal tax $\tau(\theta) := p(\theta) - c$ that each consumer faces. As Diamond (1973) showed, the optimal constant marginal tax is equal to a weighted average of the externality generated by each consumer (cf. the discussion following Corollary 1). However, as Proposition 3 clarifies, this formula is simply equal to the projection of the externality ξ onto the set of constant marginal taxes. More generally, Proposition 3 implies that the optimal marginal tax in any set of marginal taxes is a projection of ξ onto that set.

Finally, the marginal price that implements the second-best allocation can be derived by combining Proposition 3 with Theorem 2. On one hand, when the second-best allocation is the same for any two consumers, they must face the same marginal price. On the other hand, notice that the Bregman divergence must obey the "law of cosines":

$$\begin{aligned}\Delta_{\text{CS}}^{\theta}\left(c+\xi,p(\theta)\right) &= \Delta_{\text{CS}}^{\theta}\left(c+\xi,c+\mathbf{E}\left[\xi \,|\, \theta\right]\right)+\Delta_{\text{CS}}^{\theta}\left(c+\mathbf{E}\left[\xi \,|\, \theta\right],p(\theta)\right)\\ &\quad+\left(\xi-\mathbf{E}\left[\xi \,|\, \theta\right]\right)\left[(v')^{-1}\left(\frac{p(\theta)}{\theta}\right)-(v')^{-1}\left(\frac{c+\mathbf{E}\left[\xi \,|\, \theta\right]}{\theta}\right)\right].\end{aligned}$$

Taking expectation with respect to $\xi \,|\, \theta$ yields the identity

$$\underbrace{\mathbf{E}\left[\Delta_{\text{CS}}^{\theta}\left(c+\xi,p(\theta)\right)\,|\,\theta\right]}_{=\mathbf{E}[\text{DWL}(\theta,\xi)\,|\,\theta]}=\mathbf{E}\left[\Delta_{\text{CS}}^{\theta}\left(c+\xi,c+\mathbf{E}\left[\xi \,|\, \theta\right]\right)\,|\,\theta\right]+\Delta_{\text{CS}}^{\theta}\left(c+\mathbf{E}\left[\xi \,|\, \theta\right],p(\theta)\right).$$

Therefore, deadweight loss is minimized by choosing $p(\theta)$ to minimize the right-hand side of this identity whenever possible—that is, by setting $p(\theta) = c + \mathbf{E}\left[\xi \,|\, \theta\right]$ whenever the second-best allocation function is not constant at θ. The following corollary summarizes this derivation.

Corollary 2. *Let $\overline{q}^{\text{FB}}$ and q^{SB} be the ironed first-best allocation function and the second-best allocation function as defined in Theorem 2. Then the marginal price function p^{SB} that implements the second-best mechanism is*

$$p^{\text{SB}}(\theta)=c+\mathbf{E}\left[\xi \,|\, \hat{\theta}(\theta)\right],\qquad \textit{where } \hat{\theta}(\theta):=\inf_{z\in[\underline{\theta},\overline{\theta}]}\left\{q^{\text{SB}}(\theta)=q^{\text{SB}}(z):q^{\text{SB}}(z)=\overline{q}^{\text{FB}}(z)\right\}.$$

5.2 Welfare analysis

The regression approach used to derive the optimal allocation and optimal marginal price also yields an intuitive decomposition of the deadweight loss due to the second-best nonlinear tax and the additional deadweight loss due to the optimal linear tax: they are, respectively, the "variance of ξ unexplained by θ" and the "variance of ξ explained by θ." For ease of exposition, I first focus on the case where the second-best allocation function is strictly increasing, and then discuss the case where it is not necessarily so at the end of this subsection.

This decomposition is easiest to see for the special case of linear demand under Assumption 1. In this case, the deadweight losses due to the second-best nonlinear tax and the optimal linear tax can be explicitly computed using Proposition 1:

$$\begin{cases}\text{DWL}^{\text{SB}} &= \dfrac{1}{2}\,\mathbf{E}\left[\dfrac{1}{\theta}\left[\xi-\mathbf{E}\left[\xi \,|\, \theta\right]\right]^2\right],\\[2ex] \text{DWL}^{\text{L}} &= \dfrac{1}{2}\,\mathbf{E}\left[\dfrac{1}{\theta}\left[\xi-\dfrac{\mathbf{E}\left[\xi/\theta\right]}{\mathbf{E}\left[1/\theta\right]}\right]^2\right]\end{cases}\implies \text{DWL}^{\text{L}}-\text{DWL}^{\text{SB}}=\frac{1}{2}\,\mathbf{E}\left[\frac{1}{\theta}\left[\mathbf{E}\left[\xi \,|\, \theta\right]-\frac{\mathbf{E}\left[\xi/\theta\right]}{\mathbf{E}\left[1/\theta\right]}\right]^2\right].$$

To interpret these expressions, consider the expectation and variance operators $\widetilde{\mathbf{E}}$ and $\widetilde{\mathbf{Var}}$ defined with respect to the distribution on $[\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$ with density proportional to $g(\theta, \xi)/\theta$. Then the deadweight loss due to the optimal linear tax can be decomposed in accordance to the law of total variance:

$$\begin{aligned}
\underbrace{\frac{1}{2}\mathbf{E}\left[\frac{1}{\theta}\left[\xi - \frac{\mathbf{E}[\xi/\theta]}{\mathbf{E}[1/\theta]}\right]^2\right]}_{\mathrm{DWL^L}} &= \frac{1}{2}\widetilde{\mathbf{E}}\left[\left(\xi - \widetilde{\mathbf{E}}[\xi]\right)^2\right] = \frac{1}{2}\widetilde{\mathbf{Var}}[\xi] \\
&= \frac{1}{2}\widetilde{\mathbf{E}}\left[\widetilde{\mathbf{Var}}[\xi \mid \theta]\right] + \frac{1}{2}\widetilde{\mathbf{Var}}\left[\widetilde{\mathbf{E}}[\xi \mid \theta]\right] \\
&= \underbrace{\frac{1}{2}\mathbf{E}\left[\frac{1}{\theta}\mathbf{E}\left[[\xi - \mathbf{E}[\xi \mid \theta]]^2 \mid \theta\right]\right]}_{\mathrm{DWL^{SB}}} + \underbrace{\frac{1}{2}\mathbf{E}\left[\frac{1}{\theta}\left[\mathbf{E}[\xi \mid \theta] - \frac{\mathbf{E}[\xi/\theta]}{\mathbf{E}[1/\theta]}\right]^2\right]}_{\mathrm{DWL^L} - \mathrm{DWL^{SB}}}.
\end{aligned}$$

This shows that the deadweight loss due to the optimal linear tax can be interpreted as the variance of ξ. By using the second-best nonlinear tax instead, the social planner can achieve a welfare gain equal to the variance of ξ that is explained by θ. However, a deadweight loss equal to the variance of ξ that is unexplained by θ remains: this deadweight loss arises because the social planner cannot measure and directly tax the externality.

This decomposition can be extended to the case of general demand by defining suitable analogs of the expectation, variance, and conditional variance operators. For any function $\psi : [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$, define the analog $\mathcal{E}$ of the expectation operator by

$$\mathcal{E}[\psi(\theta, \xi)] \in \arg\min_{\tau \in \mathbb{R}} \mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}(c + \psi(\theta, \xi), c + \tau)\right].$$

This analogy is justified in the special case where the distribution of θ is degenerate, in which case $\mathcal{E}[\psi(\theta, \xi)] = \mathbf{E}[\psi(\theta, \xi)]$ (Banerjee, Merugu, Dhillon, Ghosh, and Lafferty, 2005). Further define the analog $\mathcal{V}$ of the variance operator by

$$\mathcal{V}[\psi(\theta, \xi)] := \mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}(c + \psi(\theta, \xi), c + \mathcal{E}[\psi(\theta, \xi)])\right].$$

Similar to the variance operator, $\mathcal{V}$ measures the "average distance" between $\psi(\theta, \xi)$ and $\mathcal{E}[\psi(\theta, \xi)]$. Finally, the analog $\mathcal{V}[\cdot \mid \theta]$ to the conditional variance operator is defined by

$$\mathcal{V}[\psi(\theta, \xi) \mid \theta] := \mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}(c + \psi(\theta, \xi), c + \mathbf{E}[\psi(\theta, \xi) \mid \theta])\right].$$

Similar to the conditional variance operator, $\mathcal{V}[\cdot \mid \theta]$ measures the residual variance from using $\mathbf{E}[\psi(\theta,\xi) \mid \theta]$ to predict $\psi(\theta,\xi)$. Using these definitions, the following proposition derives an analog of the law of total variance, which extends the deadweight loss decomposition to the case of general demand.

Proposition 4. *The deadweight loss due to the optimal linear tax is $\mathcal{V}[\xi]$, which satisfies the identity*

$$\mathcal{V}[\xi] = \mathbf{E}[\mathcal{V}[\xi \mid \theta]] + \mathcal{V}[\mathbf{E}[\xi \mid \theta]].$$

When the second-best allocation function is strictly increasing, $\mathcal{V}[\xi \mid \theta]$ is equal to the deadweight loss due to the second-best nonlinear tax, and $\mathcal{V}[\mathbf{E}[\xi \mid \theta]]$ is equal to the additional deadweight loss due to the restriction to a linear tax.

Proposition 4 is summarized by Figure 4.

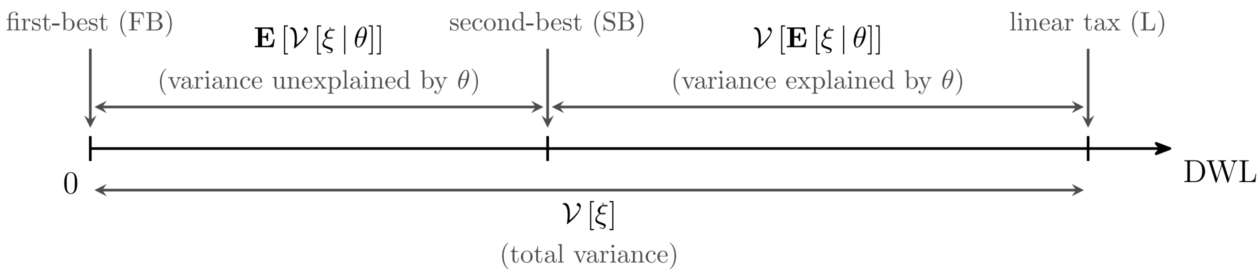

Figure 4: Decomposition of deadweight loss in Proposition 4.

The deadweight loss decomposition given by Proposition 4 has implications for policy. While linear taxes might often be easier to implement, significant welfare gain can sometimes be achieved by nonlinear taxes. Proposition 4 provides a way to evaluate how much welfare is lost by restriction to a linear tax via the fraction of variance unexplained, $\mathcal{V}[\mathbf{E}[\xi \mid \theta]] / \mathcal{V}[\xi]$. The closer this fraction is to 1, the more the policymaker should consider using nonlinear taxes.

Finally, when the second-best allocation function is not necessarily strictly increasing, the deadweight loss due to the second-best nonlinear tax will generally be larger than $\mathbf{E}[\mathcal{V}[\xi \mid \theta]]$. An additional deadweight loss arises because of the binding incentive-compatibility constraint (IC). In this case, the fraction of variance unexplained nevertheless provides an upper bound on how much deadweight loss due to the optimal linear tax can be overcome with a nonlinear tax. In the next section, I examine this case and establish necessary and sufficient conditions under which the second-best allocation function requires certain quantity restrictions.

5.3 When are quantity restrictions optimal?

I now examine when quantity restrictions are optimal. I begin by characterizing when extreme forms of these restrictions—namely, bans and mandates—are justified.

Proposition 5 (bans and mandates)**.**

(i) *The second-best mechanism is a ban (i.e.,* $q^{\mathrm{SB}}(\theta) \equiv 0$*) if and only if*

$$\int_{\hat{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} [c + \xi - \theta v'(0)] \ \mathrm{d}G(\theta, \xi) \geq 0 \qquad \textit{for any } \hat{\theta} \in [\underline{\theta}, \overline{\theta}].$$

(ii) *The second-best mechanism is a mandate (i.e.,* $q^{\mathrm{SB}}(\theta) \equiv A$*) if and only if*

$$\int_{\underline{\theta}}^{\hat{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} [\theta v'(A) - c - \xi] \ \mathrm{d}G(\theta, \xi) \geq 0 \qquad \textit{for any } \hat{\theta} \in [\underline{\theta}, \overline{\theta}].$$

Proposition 5 gives simple necessary and sufficient conditions under which bans or mandates are optimal. In the case of a ban, the expected social marginal cost must intuitively be sufficiently high so as to outweigh the expected marginal benefit from consuming any quantity of the good. This intuition is formalized by the first condition in the proposition, which states that conditional on having a consumption type above any $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, the expected social marginal cost must be no less than the expected marginal benefit from consuming the zeroth unit of the good. In the case of a mandate, it is more natural to think of $\xi < 0$ so that consumption of the good generates a positive externality, such as a vaccine that prevents the spread of a virus. Nevertheless, a similar condition holds, as formalized by the second condition in the proposition: conditional on having a consumption type above any $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, the expected marginal benefit must be no less than the expected social marginal cost from consuming the zeroth unit of the good.

To connect these conditions to the motivating example of a vaccine, which has unit demand, a technical observation should be noted here. Although I have focused on the case of continuous demand, it should be pointed out that Proposition 5 (and, more generally, the rest of my analysis) extends straightforwardly to the case of unit demand, such as for a vaccine. To do so, set $A = 1$, consider the sequence of functions $v_\varepsilon(q) = q - \varepsilon q^2/2$ for $0 < \varepsilon < 1$, and take the limit $\varepsilon \to 0$. In this limit, the quantity of the good consumed by a consumer can be interpreted as the probability that he receives the good. Because each v_ε is an increasing and strictly concave function, it satisfies the assumptions of the model presented in Section 2. As such, Proposition 5 applies to each v_ε.

Since $v'_\varepsilon(q) \to 1$ pointwise, it follows that the second-best mechanism is a mandate in the limiting unit demand model if and only if $\mathbf{E}\left[\theta - c - \xi \,|\, \theta \geq \hat{\theta}\right] \geq 0$ for any $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$.

Proposition 5 therefore provides support for the use of bans and mandates to indirectly regulate externalities. In reality, the conditions stated in the proposition might even understate the actual use of bans and mandates for at least two reasons. First, policymakers might find it difficult to implement complicated nonlinear taxes in some settings. Therefore, even if a ban or mandate is not second-best, it might be nonetheless optimal within the more constrained set of policies that the policymaker can easily implement. Second, quantity regulations are often cheaper to enforce than taxes due to the ease of identifying violations (Glaeser and Shleifer, 2001); an analysis that accounts for enforcement costs might thus favor the use of bans and mandates even more.

Even if bans and mandates are not optimal, quantity restrictions can nonetheless be justified. Consider two types of quantity restrictions: quantity floors and quantity ceilings. These impose a minimum quantity and a maximum quantity of consumption respectively. Since any mechanism can stipulate a non-binding minimum or maximum quantity, I say that a mechanism (q, t) imposes a *quantity floor* only if the specified minimum quantity binds for a positive measure of consumers: that is, if $q(\theta) = \underline{q} = \inf \operatorname{im} q$ holds for a positive measure of consumption types θ. Similarly, a mechanism imposes a *quantity ceiling* only if the specified maximum quantity binds for a positive measure of consumers.

Proposition 6 (quantity floors and ceilings)**.**

(i) *The second-best mechanism imposes a quantity floor if and only if there exists* $\theta^* \in (\underline{\theta}, \overline{\theta}]$ *such that*

$$\frac{c + \mathbf{E}\left[\xi \,|\, \theta \leq \theta^*\right]}{c + \mathbf{E}\left[\xi \,|\, \theta = \underline{\theta}\right]} \geq \mathbf{E}\left[\frac{\theta}{\underline{\theta}} \,|\, \theta \leq \theta^*\right].$$

(ii) *The second-best mechanism imposes a quantity ceiling if and only if there exists* $\theta^* \in [\underline{\theta}, \overline{\theta})$ *such that*

$$\frac{c + \mathbf{E}\left[\xi \,|\, \theta > \theta^*\right]}{c + \mathbf{E}\left[\xi \,|\, \theta = \overline{\theta}\right]} \leq \mathbf{E}\left[\frac{\theta}{\overline{\theta}} \,|\, \theta > \theta^*\right].$$

Similar to Proposition 5, Proposition 6 gives necessary and sufficient conditions under which quantity floors and quantity ceilings are optimal. In the case of quantity floors, the first condition in Proposition 6 evaluates the ratio of the expected social marginal cost, conditional on having a consumption type not exceeding θ^*, to the expected social marginal cost, conditional on having a consumption type of $\underline{\theta}$. This ratio is compared to the ratio of the expected consumption type, conditional on having a consumption type not exceeding θ^*, to the lowest consumption type of $\underline{\theta}$.

In the case of quantity ceilings, the second condition in Proposition 6 takes a similar form, except that expectations are taken conditional on either having a consumption type exceeding θ^* or equal to $\overline{\theta}$ instead.

Intuitively, these conditions thus express that quantity floors and quantity ceilings are optimal when the ratio of the social marginal cost to the consumption type differs locally at either $\underline{\theta}$ or $\overline{\theta}$ from an average of these ratios in their neighborhoods. Then the social planner would prefer to induce a higher allocation for the lowest consumption type $\underline{\theta}$ (in the case of a quantity floor) and a lower allocation for the highest consumption type $\overline{\theta}$ (in the case of a quantity ceiling) than other consumers with neighboring consumption types. Because of incentive compatibility constraints, the social planner is unable to do so without also changing the allocations of other consumers with neighboring consumption types. Under the conditions stated in Proposition 6, the gains in doing so are just large enough that the social planner would optimal equalize the allocations of consumers with consumption types around either $\underline{\theta}$ or $\overline{\theta}$ by imposing a quantity floor or quantity ceiling.

The following corollary gives sufficient conditions for the use of quantity floors and quantity ceilings directly in terms of the rate of change of this ratio.

Corollary 3 (quantity floors and ceilings)**.** *Let $\chi(\theta)$ denote the ratio of the expected social marginal cost to the consumption type at θ:*

$$\chi(\theta) := \frac{c + \mathbf{E}\left[\xi \mid \theta\right]}{\theta}.$$

(i) *The second-best mechanism imposes a quantity floor if $\chi'(\underline{\theta}) \geq 0$.*

(ii) *The second-best mechanism imposes a quantity ceiling if $\chi'(\overline{\theta}) \geq 0$.*

6 Empirical application

I now apply my regression approach to the taxation of vehicle miles traveled (VMT) to regulate automobile externalities.

6.1 Setting

Although many automobile externalities are currently regulated via a gas tax (Parry, Walls, and Harrington, 2007), it is likely that this will change in the near future. There are at least two reasons for this change. First, advancements in telematics technology now allow cumulative miles driven to be tracked on the level of individual vehicles. This opens a new possibility for policymakers: nonlinear taxation of VMT. Whereas it would be infeasible to tax gas nonlinearly due to the potential for fraud (as pointed out by Diamond, 1973), telematics sensors that track VMT on the level of individual vehicles will eliminate this susceptibility. Second, gas vehicles are increasingly being replaced by electric vehicles. This means that the gas tax will no longer be able to indirectly regulate externalities such as accidents and road damage imposed by electric vehicles. These two reasons motivated a $125 million pilot study for a VMT tax as part of the Bipartisan Infrastructure Bill, which was signed into law by U.S. President Joe Biden in November 2021. However, even if the rapid uptake of electric vehicles was set aside, a nonlinear VMT tax promises to be more efficient than a linear gas tax simply because the indirect tax is no longer constrained to be linear (cf. Proposition 4).

To evaluate the potential gain in efficiency of this policy change, a key question concerns the empirical relationship between automobile externalities and VMT consumption. In the context of my model (cf. Section 2), a key determinant of the optimal indirect policy is the joint distribution G of VMT consumption preferences, θ, and externalities generated by VMT consumption, ξ. In particular, the correlation between θ and ξ determines how much welfare gain a nonlinear VMT tax might potentially achieve over a linear VMT tax or a linear gas tax.

I therefore examine how the relationship between emission externalities and VMT consumption affects the design of the optimal VMT tax. To this end, I apply my regression approach to the indirect regulation of CO, HCs, and NO_x pollutant emissions, following the study of Knittel and Sandler (2018). Using data from the Californian Smog Check Program between 1998 and 2008, Knittel and Sandler estimate the price elasticity of VMT demand and show that it is correlated with the amount of externality produced by each vehicle. They then compute the optimal linear gas tax and find that a substantial deadweight loss would remain even if the optimal linear gas tax was to be levied. Below, I show that the performance of the optimal linear VMT tax is roughly comparable to that of the optimal linear gas tax; however, I also show that the optimal nonlinear VMT tax performs substantially better.

To estimate the relationship between emission externalities and VMT consumption, Knittel and Sandler posit a linear demand model identical to the one presented in Section 3. Consumption preferences for VMT are determined by a consumption type θ, which is equal in magnitude to the slope of the consumer's inverse demand curve. The social marginal cost ξ of VMT consumption is constant, but it differs across consumers due to the heterogeneity in CO, HCs, and NO_x pollutants generated per mile.

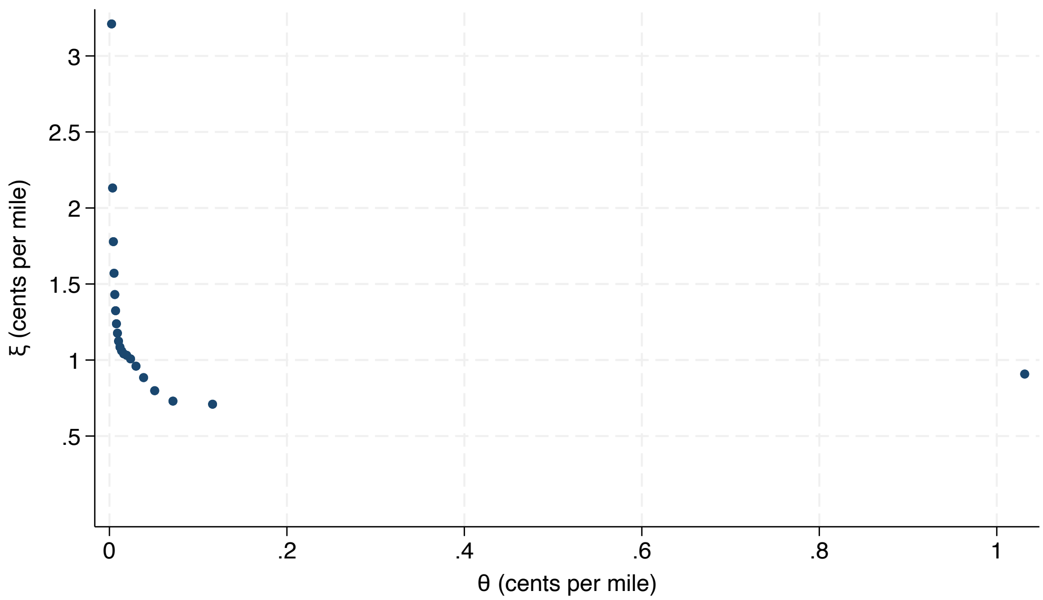

Figure 5: Binned scatter plot of empirical relationship between externality and consumption types.

Figure 5 plots the relationship between θ and ξ as implied by Knittel and Sandler's estimates for all years in the sample. Clearly, θ and ξ are negatively correlated. Since θ is equal in magnitude to the slope of the consumer's inverse demand curve, it measures the price sensitivity of the consumer: consumers with a lower θ are more price-sensitive. Therefore, Figure 5 shows that consumers who are more price-sensitive also tend to generate a greater externality per mile.

6.2 Results

Using my regression approach, I compute the deadweight loss due to three VMT taxes: the optimal linear VMT tax, a simple nonlinear VMT tax which I describe below, and the optimal nonlinear (i.e., second-best) VMT tax. I compare these to the deadweight loss due to the optimal linear gas tax as computed by Knittel and Sandler. The results are shown in Table 1.

As Table 1 shows, a substantial deadweight loss remains even after imposing the optimal linear VMT tax. Given the relationship between θ and ξ as shown in Figure 5, this observation is not surprising. As Proposition 3 implies, the deadweight loss due to the optimal linear tax is equal to the residual sum of squares from a projection of ξ onto a *constant* function of θ, since the marginal tax is constant. However, the relationship between θ and ξ as shown in Figure 5 is obviously not well described by a constant function.

Year	DWL			
	linear gas	linear VMT	simple nonlinear VMT	second-best VMT
1998	\$111.6 M	\$86.3 M	\$81.2 M	\$74.5 M
1999	\$91.2 M	\$70.2 M	\$66.1 M	\$60.2 M
2000	\$76.5 M	\$59.6 M	\$56.1 M	\$51.2 M
2001	\$63.0 M	\$51.4 M	\$48.8 M	\$44.8 M
2002	\$51.8 M	\$43.1 M	\$41.3 M	\$38.0 M
2003	\$41.2 M	\$34.2 M	\$32.9 M	\$30.6 M
2004	\$31.5 M	\$26.8 M	\$26.1 M	\$24.6 M
2005	\$20.9 M	\$18.3 M	\$18.0 M	\$17.3 M
2006	\$16.5 M	\$14.8 M	\$14.6 M	\$14.1 M
2007	\$12.7 M	\$11.5 M	\$11.4 M	\$11.0 M
2008	\$10.3 M	\$9.5 M	\$9.4 M	\$9.1 M
Average	\$47.9 M	\$38.7 M	\$36.9 M	\$34.1 M

Table 1: Deadweight loss by year under different counterfactual policies.

Table 1 also shows that the deadweight loss due to the optimal linear VMT tax yields a smaller deadweight loss relative to the optimal linear gas tax. This is because a tax on gas is exposed to heterogeneity in fuel efficiency, which introduces more variance in externality and dampens the correlation between externality and consumption as now gas consumption, rather than VMT consumption, is taxed. In the context of Proposition 4, this improves efficiency of the indirect policy since it reduces the variance $\mathcal{V}[\xi]$ of the externality.

However, the policymaker can generally do better than the optimal linear VMT tax, as implied by Theorem 2. The decreasing relationship between ξ and θ as shown in Figure 5 suggests the use of different marginal taxes for consumers with low θ and for consumers with high θ. To illustrate, consider a simple nonlinear VMT tax that sets a constant marginal tax rate for consumers with an above-median θ and a different constant marginal tax rate for consumers with a below-median θ. The deadweight loss due to this simple nonlinear VMT tax is shown in the fourth column of Table 1; on average, despite its simplicity, it performs 3.5% better than the optimal linear VMT tax and entails an annual welfare gain of \$1.8M.

Using Corollary 2, I determine the second-best VMT tax by non-parametrically regressing ξ on θ. By Proposition 3 and Proposition 4, the R^2 of this weighted least squares regression gives the fraction of the deadweight loss due to the optimal linear VMT tax that can be recovered. I use this to compute the last column of Table 1. On average, the second-best VMT tax performs 9.5% better than the optimal linear VMT tax and entails an annual welfare gain of \$2.8M.

Table 1 also shows that the amount of deadweight loss that can be recovered by any of the indirect policies decreases over time. This arises for two reasons. First, the total deadweight loss decreases over the sample period as the fleet of vehicles in California became cleaner. As Knittel and Sandler observe, one reason for this is that the U.S. Environmental Protection Agency tightened limits on the emissions per mile of new vehicles, whose overall share in the fleet grew over time. California also has more stringent new-vehicle emissions standards than other U.S. states, which contributed to the shrinking of the total deadweight loss. With a smaller total deadweight loss, the deadweight loss due to a linear gas tax or a linear VMT tax (i.e., the first and second columns in Table 1) would be expected to decrease over time. Second, as Knittel and Sandler show, the skew of the distribution of externalities also reduced over this period. This could have the effect of dampening the correlation between ξ and θ, which would leave a greater share of the variance of ξ unexplained by θ. As Proposition 3 implies, this would contribute to the reduced effectiveness of nonlinear taxation over time, as shown in the third and fourth columns of Table 1.

Overall, this application demonstrates how my regression approach can be applied to empirical settings. As I have shown, the optimal indirect policy in each of three families of policies—linear VMT taxes, simple 2-piece piecewise-linear VMT taxes, and nonlinear VMT taxes—can be computed by estimating a different regression. The welfare implications of each policy can be easily computed via an analysis of variance for each of the estimated regressions.

7 Extensions

In this section, I discuss how my regression approach and corresponding results extend in two ways: when the externality takes a more general form and when the social planner has revenue and/or distributional concerns.

7.1 More general externality

While I have assumed that consumers' utility functions are additively separable in the externality for ease of exposition, extending the approach to more general utility functions is straightforward. Suppose that, if a consumer with type (θ, ξ) consumes a quantity $q(\theta, \xi)$ of the good and makes a payment of $t(\theta, \xi)$, then he realizes a utility of

$$\theta v(q(\theta,\xi), E) - t(\theta,\xi), \quad \text{where } E = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \xi q(\theta', \xi') \; \mathrm{d}G(\theta', \xi').$$

This is the case, for example, if the externality generated by other consumers reduces the benefit that the consumer in question derives from the good. For instance, congestion might reduce the consumption benefit that each consumer derives from driving. I assume that v is continuous in both arguments, and that $v(\cdot, E)$ is twice continuously differentiable, increasing, and concave for any value $E \in \mathbb{R}$ of the total externality.

The only way in which a non-additively separable externality changes the analysis is that the social marginal cost of the externality might generally depend on the total amount of externality. The social planner solves:

$$\max_{q,E} \left\{ \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \left[\theta v(q(\theta), E) - cq(\theta)\right] \, \mathrm{d}G(\theta, \xi) : \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \xi q(\theta) \, \mathrm{d}G(\theta, \xi) = E \right\}.$$

For any given value of E, the constraint on total externality can be incorporated into the social planner's objective through a Lagrangian multiplier λ, which represents the shadow cost of the externality when the total externality is E. Consequently, the approach developed in Section 3.2 nonetheless applies for that value of E with a social marginal cost of $c + \lambda\xi$.

A different situation with a more general form of externality arises when consumers generate an externality that varies nonlinearly with consumption. Suppose that a consumer with an externality type of ξ generate $\xi e(q)$ units of the externality when they consume q units of the good, where $e : \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing convex function. This nests the model presented in Section 2 as a special case (i.e., when $e(q) = q$). Then the regression approach developed in Section 4 still holds with a change of variables: by letting $x(\theta) = e(q(\theta))$ and defining the increasing concave function $\widetilde{v} = v \circ e^{-1}$, the previous analysis goes through by replacing q with x and v with $\widetilde{v}$. Qualitatively, however, any nonlinearity in e will favor a nonlinear tax. Indeed, even the first-best benchmark in this case can generally only be implemented by a nonlinear tax on the good.

7.2 Revenue and distributional concerns

In many settings of interest—alcohol, cigarette, and gas taxes—policymakers might have revenue concerns in addition to the objective of regulating externalities. Suppose that the social planner places some weight $\alpha \in \mathbb{R}_+$ on the total revenue that the tax raises, so that she solves:

$$\max_{q} \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \left\{ \left[\theta - \alpha \frac{1 - F(\theta)}{f(\theta)}\right] v(q(\theta)) - (c + \xi)\, q(\theta) \right\} \, \mathrm{d}G(\theta, \xi), \tag{2}$$

where F denotes the marginal distribution of θ with density f on the interval $[\underline{\theta}, \overline{\theta}]$, induced by the joint distribution G. For ease of notation, write $J_\alpha(\theta) = \theta - \alpha\,[1 - F(\theta)]\,/f(\theta)$; $\alpha = 0$ corresponds to the case of a social planner who maximizes total social surplus, while $\alpha = 1$ corresponds to the case of a social planner who maximizes total revenue.

Although the social planner has a more general objective function, a regression approach nonetheless applies with a change of variables. By setting $x(\theta) = v(q(\theta))$, write the social planner's problem as

$$\min_x \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} (c+\xi)\left[v^{-1}(x(\theta)) - \frac{J_\alpha(\theta)}{c+\xi}x(\theta)\right]\,\mathrm{d}G(\theta,\xi).$$

Since v is increasing and concave, hence v^{-1} is convex. It follows that the social planner's objective can again be written as a weighted sum of Bregman divergences.

However, there are two notable differences between this approach and the one developed in Section 4. First, instead of θ, the Bregman divergences here are weighted by the social marginal cost $c + \xi$. This difference arises because $J_\alpha(\theta)$, unlike θ, is not guaranteed to be positive when $\alpha \neq 0$. Intuitively, because uniform transfers from consumers to the social planner are no longer welfare neutral, the social planner must now account for the difference in utilities arising from transfers; hence, the regression that the social planner performs must be in utilities rather than in quantities. Second, instead of the first-best allocation function, the solution to the social planner's problem here is equal to the projection of the "target allocation function,"

$$q^{\mathrm{T}}(\theta,\xi) = (v')^{-1}\left(\frac{c+\xi}{J_\alpha(\theta)}\right).$$

This target allocation function is what the social planner would choose if the externality could be measured and directly taxed. The target allocation function is distorted downwards relative to the first-best allocation function: the social planner sets a higher tax to extract higher tax revenue, which induces inefficiently lower consumption. It can also be shown using the regression approach that a similar comparative static holds for the solution to the social planner's problem (i.e., the projection of the target allocation function onto the set of implementable allocations).

A different but related concern that can arise is the distributional impact of corrective taxation: the policymaker might face an additional trade-off between equity and efficiency.[5] Suppose that the social planner assigns an expected weight of $\omega(\theta)$ to the utility of a consumer with consumption type θ. A standard envelope theorem argument shows that the social planner's problem can be

[5] Akbarpour ⓡ Budish ⓡ Dworczak ⓡ Kominers (2023) and Pai and Strack (2022) analyze how the trade-off between equity and efficiency affects the direct taxation of externalities.

written in a similar form as equation (2), which can be solved in an identical fashion to the case of revenue concerns, albeit with a different target allocation function.

8 Conclusion

Policymakers often introduce quantity controls to regulate goods that cause externalities because the amount of externalities generated by each consumer is heterogeneous and difficult to measure. This paper provides some justification for these policies by showing that quantity controls can be necessary in the optimal policy. This is because the amount of the externality generated can be correlated with the amount of the good demanded, even if the policymaker can tax only the good. As I show, quantity controls—and, more generally, nonlinear taxes—can exploit this correlation so as to minimize the deadweight loss of indirect taxation.

To analyze the deadweight loss of indirect taxation, this paper develops an approach based on regression. This approach is motivated by the classic analysis of Harberger (1964), which is widely expanded on in the sufficient statistics literature (Chetty, 2009; Kleven, 2021). While Harberger's original analysis applies only to linear taxes in the case of linear demand, this paper's contribution is to show that a combination of Harberger's analysis and mechanism design techniques produces a unified approach for nonlinear taxes and general demand systems. This approach shows that the deadweight loss of indirect taxation is equal to the residual of a regression of the first-best allocation function onto any given family of allocation functions. The optimal allocation function in that family is thus equal to the corresponding projection of the the first-best allocation function. Using this approach, I derive the second-best policy by characterizing the projection operator onto the family of implementable allocation functions and show that the second-best allocation function can require quantity controls. This can occur, for example, when consumers who have the highest demand tend to generate disproportionately more external harm, resulting in the optimal use of a quantity ceiling to cap the total amount of external harm generated.

In addition to characterizing optimal policies, my approach also uncovers new implications for indirect taxation. I illustrate with three insights. First, the marginal prices that implement the optimal allocation can be derived by performing the regression in prices rather than quantities, which allows for a reinterpretation and generalization of Diamond's (1973) classic formula for the optimal linear tax. Second, applying the law of total variance to this regression sheds light on welfare: whereas the deadweight loss of the optimal linear indirect tax arises due to the variance in externality, the second-best nonlinear indirect tax is able to recover up to the variance in externality that is explained by consumption. Third, this regression approach allows the optimal indirect tax to be estimated empirically, which I demonstrate by applying my results to the taxation of vehicle miles traveled for regulating automobile externalities.

The usefulness of this approach likely extends beyond indirect taxation in the regulation of externalities. On one hand, this approach might be able to shed new light on the theory of optimal pricing and regulation in settings with internalities and/or adverse selection, such as in the design of health insurance policies. On the other hand, the widespread use of Harberger triangles in empirical work highlights the potential for the approach developed by this paper to analyze optimal nonlinear taxation and counterfactual welfare gains in the same empirical settings. In particular, settings with imperfect pricing policies (Jacobsen et al., 2020)—such as electricity pricing with wedges between marginal cost of production and real-time pricing, durable goods with heterogeneity in utilization, and congestion pricing with imperfect spatial and temporal policy differentiation—seem amenable to study with this approach.

References

Akbarpour, M. ⓡ E. Budish ⓡ P. Dworczak ⓡ S. D. Kominers (2023): "An Economic Framework for Vaccine Prioritization," *Working paper.*

Akbarpour, M. ⓡ P. Dworczak ⓡ S. D. Kominers (2022): "Redistributive Allocation Mechanisms," *Working paper.*

Balakrishnan, A. V. (1981): *Applied Functional Analysis*, New York, NY: Springer-Verlag, 2nd edition.

Banerjee, A., S. Merugu, I. S. Dhillon, J. Ghosh, and J. Lafferty (2005): "Clustering with Bregman Divergences.," *Journal of Machine Learning Research*, 6.

Bovenberg, A. L. and F. van der Ploeg (1994): "Environmental Policy, Public Finance and the Labour Market in a Second-Best World," *Journal of Public Economics*, 55, 349–390.

Che, Y.-K., W. Dessein, and N. Kartik (2013): "Pandering to Persuade," *American Economic Review*, 103, 47–79.

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

Condorelli, D. (2013): "Market and Non-Market Mechanisms for the Optimal Allocation of Scarce Resources," *Games and Economic Behavior*, 82, 582–591.

Cremer, H., F. Gahvari, and N. Ladoux (1998): "Externalities and Optimal Taxation," *Journal of Public Economics*, 70, 343–364.

Diamond, P. A. (1973): "Consumption Externalities and Imperfect Corrective Pricing," *Bell Journal of Economics and Management Science*, 526–538.

Dworczak, P. ⓡ S. D. Kominers ⓡ M. Akbarpour (2021): "Redistribution Through Markets," *Econometrica*, 89, 1665–1698.

Fleischer, V. (2015): "Curb Your Enthusiasm for Pigovian Taxes," *Vanderbilt Law Review*, 68, 1673–1713.

Glaeser, E. L. and A. Shleifer (2001): "A Reason for Quantity Regulation," *American Economic Review (Papers and Proceedings)*, 91, 431–435.

Griffith, R., M. O'Connell, and K. Smith (2019): "Tax Design in the Alcohol Market," *Journal of Public Economics*, 172, 20–35.

Grummon, A. H., B. B. Lockwood, D. Taubinsky, and H. Allcott (2019): "Designing Better Sugary Drink Taxes," *Science*, 365, 989–990.

Harberger, A. C. (1964): "The Measurement of Waste," *American Economic Review (Papers and Proceedings)*, 54, 58–76.

Jacobsen, M. R., C. R. Knittel, J. M. Sallee, and A. A. van Benthem (2020): "The Use of Regression Statistics to Analyze Imperfect Pricing Policies," *Journal of Political Economy*, 128, 1826–1876.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2021): *An Introduction to Statistical Learning with Applications in R*, New York, NY: Springer, 2nd edition.

Jehiel, P. and B. Moldovanu (2001): "Efficient Design with Interdependent Valuations," *Econometrica*, 69, 1237–1259.

Kang, Z. Y. (2023): "The Public Option and Optimal Redistribution," *Working paper*.

Kaplow, L. (2012): "Optimal Control of Externalities in the Presence of Income Taxation," *International Economic Review*, 53, 487–509.

Kaplow, L. and S. Shavell (2002): "On the Superiority of Corrective Taxes to Quantity Regulation," *American Law and Economics Review*, 4, 1–17.

Kleven, H. J. (2021): "Sufficient Statistics Revisited," *Annual Review of Economics*, 13, 515–538.

Knittel, C. R. and R. Sandler (2018): "The Welfare Impact of Second-Best Uniform-Pigouvian Taxation: Evidence from Transportation," *American Economic Journal: Economic Policy*, 10, 211–42.

Kopczuk, W. (2003): "A Note on Optimal Taxation in the Presence of Externalities," *Economics Letters*, 80, 81–86.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Ng, Y.-K. (1980): "Optimal Corrective Taxes or Subsidies when Revenue Raising Imposes an Excess Burden," *American Economic Review*, 70, 744–751.

Pai, M. and P. Strack (2022): "Taxing Externalities Without Hurting the Poor," *Working paper.*

Parry, I. W., M. Walls, and W. Harrington (2007): "Automobile Externalities and Policies," *Journal of Economic Literature*, 45, 373–399.

Pigou, A. C. (1920): *The Economics of Welfare*, London, UK: Macmillan.

Roberts, M. J. and M. Spence (1976): "Effluent Charges and Licenses under Uncertainty," *Journal of Public Economics*, 5, 193–208.

Sandmo, A. (1975): "Optimal Taxation in the Presence of Externalities," *Swedish Journal of Economics*, 77, 86–98.

Weitzman, M. L. (1974): "Prices vs. Quantities," *Review of Economic Studies*, 41, 477–491.

Appendix A Omitted proofs

This appendix supplies omitted proofs, which are presented in chronological order of results in the main text of the paper.

A.1 Proof of Lemma 1

Since individual reports do not affect the aggregate externality in the market, observe that the incentive compatibility (IC) constraint implies that

$$(\theta, \xi) \in \underset{(\hat{\theta}, \hat{\xi})}{\arg\max} \left[\theta v(q(\hat{\theta}, \hat{\xi})) - t(\hat{\theta}, \hat{\xi}) \right].$$

Therefore, for any $(\theta, \xi), (\theta', \xi') \in [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$,

$$\begin{cases} \theta v(q(\theta, \xi)) - t(\theta, \xi) & \geq \theta v(q(\theta', \xi')) - t(\theta', \xi'), \\ \theta' v(q(\theta', \xi')) - t(\theta', \xi') & \geq \theta' v(q(\theta, \xi)) - t(\theta, \xi). \end{cases}$$

Subtracting the second expression from the first yields

$$(\theta - \theta') \left[v(q(\theta, \xi) - v(q(\theta', \xi')) \right] \quad \text{for any } \theta, \xi), (\theta', \xi') \in [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}].$$

This implies that

$$q(\theta, \xi) \geq q(\theta', \xi') \quad \text{for any } \theta \geq \theta' \text{ and } \xi, \xi' \in [\underline{\xi}, \overline{\xi}].$$

In turn, this means that, for sufficiently small ε,

$$q(\theta + \varepsilon, \xi) \geq q(\theta, \xi') \quad \text{for almost any } \theta \in [\underline{\theta}, \overline{\theta}] \text{ and } \xi, \xi' \in [\underline{\xi}, \overline{\xi}].$$

Taking the limit $\varepsilon \to 0$ yields

$$q(\theta, \xi) \geq q(\theta, \xi') \quad \text{for almost any } \theta \in [\underline{\theta}, \overline{\theta}] \text{ and } \xi, \xi' \in [\underline{\xi}, \overline{\xi}].$$

This implies that there exists $\hat{q} \in \mathcal{Q}$ such that $q(\theta, \xi) = \hat{q}(\theta)$ for almost every $(\theta, \xi) \in [\underline{\theta}, \overline{\theta}] \times [\underline{\xi}, \overline{\xi}]$, as claimed.

A.2 Proof of Lemma 2

Let $L^2([\underline{\theta},\overline{\theta}];\mathrm{d}W)$ denote the set of functions on $[\underline{\theta},\overline{\theta}]$ that are square-integrable with respect to the measure $\mathrm{d}W$, where W has density w on $[\underline{\theta},\overline{\theta}]$. Observe that $\mathcal{Q}\subset L^2([\underline{\theta},\overline{\theta}];\mathrm{d}W)$ is nonempty, closed, and convex. Therefore, by the Hilbert projection theorem, a unique projection exists for any $q_0\in L^2([\underline{\theta},\overline{\theta}];\mathrm{d}W)$ onto $\mathcal{Q}$.

Next, to show that its unique solution is given by q^*, begin by assuming that $\overline{q}_0(\theta)\in[0,A]$ for all $\theta\in[\underline{\theta},\overline{\theta}]$. Define

$$\mathcal{I}:=\left\{q\in L^2([\underline{\theta},\overline{\theta}];\mathrm{d}W)\text{ is non-decreasing}\right\}.$$

Let $\Pi_{\mathcal{I}}(q_0)$ denote the projection of q_0 onto $\mathcal{I}$. Since $\mathcal{I}\subset L^2([\underline{\theta},\overline{\theta}];\mathrm{d}W)$ is also nonempty, closed, and convex, the projection is unique, hence $\Pi_{\mathcal{I}}(q_0)$ is well-defined. Since $\mathcal{I}$ is a convex cone,

$$\int_{\underline{\theta}}^{\overline{\theta}}\left[\Pi_{\mathcal{I}}(q_0)(\theta)-q_0(\theta)\right]q(\theta)w(\theta)\ \mathrm{d}\theta\geq 0\quad\text{for any }q\in\mathcal{I}.$$

In particular, choose $q(\theta)=\mathbf{1}_{\theta>t}$ for some $t\in[\underline{\theta},\overline{\theta}]$. Then the above implies that

$$\overline{Q}(t):=\int_t^{\overline{\theta}}\Pi_{\mathcal{I}}(q_0)(\theta)w(\theta)\ \mathrm{d}\theta\geq\int_t^{\overline{\theta}}\chi(\theta)w(\theta)\ \mathrm{d}\theta=:Q(t).$$

Since $\Pi_{\mathcal{I}}(q_0)$ is non-decreasing on $[\underline{\theta},\overline{\theta}]$, $\overline{Q}$ must be concave; hence $\overline{Q}$ is a concave majorant of Q. Now, if $\overline{Q}$ is not the *least* concave majorant of Q, then there exist $\underline{\theta}<\theta_1<\theta_2<\overline{\theta}$ and a separating linear function $\ell(\theta)$, such that $\ell(\theta)\geq Q(\theta)$ for $\theta\in[\underline{\theta},\overline{\theta}]$; $\overline{Q}(\theta)>\ell(\theta)$ for $\theta\in(\theta_1,\theta_2)$; and $\overline{Q}(\theta_i)=\ell(\theta_i)$ for $i=1,2$. Define $\widetilde{q}$ by $\widetilde{q}(\theta)=\Pi_{\mathcal{I}}(q_0)(\theta)$ for $\theta\notin(\theta_1,\theta_2)$ and $\widetilde{q}(\theta)=\ell'(\theta)$ for $\theta\in(\theta_1,\theta_2)$. Then integration by parts yields the following contradiction:

$$\begin{aligned}0&\leq\int_{\underline{\theta}}^{\overline{\theta}}\left[\Pi_{\mathcal{I}}(q_0)(\theta)-q_0(\theta)\right]\left[\widetilde{q}(\theta)-\Pi_{\mathcal{I}}(q_0)(\theta)\right]w(\theta)\ \mathrm{d}\theta\\&=-\int_{\underline{\theta}}^{\overline{\theta}}\left[Q(\theta)-\overline{Q}(\theta)\right]w(\theta)\ \mathrm{d}\left[\widetilde{q}(\theta)-\Pi_{\mathcal{I}}(q_0)(\theta)\right]=\int_{\theta_1}^{\theta_2}\left[Q(\theta)-\overline{Q}(\theta)\right]w(\theta)\mathrm{d}\Pi_{\mathcal{I}}(q_0)(\theta)<0.\end{aligned}$$

Therefore $\overline{Q}$ is the least concave majorant of Q. It follows that $\Pi_{\mathcal{I}}(q_0)=-\left(\operatorname{co}Q\right)'=\overline{q}_0$.

It remains to consider the case where either $\overline{q}_0(\theta)\leq 0$ or $\overline{q}_0(\theta)\geq A$ for some $\theta\in[\underline{\theta},\overline{\theta}]$. Define t so that $\overline{\chi}(\theta)\leq 0$ for all $\theta\in[\underline{\theta},t)$; and $0<\overline{q}_0(\theta)$ for all $t\in(t,\overline{\theta}]$. Such a $t\in[\underline{\theta},\overline{\theta}]$ exists since $\overline{q}_0$ is non-decreasing by construction. The argument above shows that $\overline{q}_0$ minimizes the integral between t and $\overline{\theta}$:

$$\overline{q}_0 \in \arg\min_{q \in \mathcal{Q}} \int_t^{\overline{\theta}} \left[q(\theta) - q_0(\theta)\right]^2 w(\theta) \, \mathrm{d}\theta.$$

Since $\mathcal{Q} \subset \mathcal{I}$, the argument above also shows that the constraint $q(\theta) \geq 0$ must bind for $\theta \in [\underline{\theta}, t]$. The case where $\overline{q}_0(\theta) \geq A$ can be similarly dealt with. This yields the solution q^* as defined in Lemma 2.

A.3 Proof of Lemma 3

Using the definition of the Bregman divergence, it can be seen that

$$\begin{aligned}
&\Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{SB}}(\theta)\right) + \Delta_{-v}^{(\theta,\xi)}\left(q^{\mathrm{SB}}(\theta), q^{\mathrm{FB}}(\theta,\xi)\right) \\
&= \Delta_{-v}^{(\theta,\xi)}\left(q(\theta), q^{\mathrm{FB}}(\theta,\xi)\right) - \left[q^{\mathrm{SB}}(\theta) - q(\theta)\right]\left[\phi^{(\theta,\xi)}(q^{\mathrm{FB}}(\theta,\xi)) - \phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta))\right].
\end{aligned}$$

Therefore, it suffices to show that, for any $q \in \mathcal{Q}$,

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \left[q^{\mathrm{SB}}(\theta) - q(\theta)\right]\left[\phi^{(\theta,\xi)}(q^{\mathrm{FB}}(\theta,\xi)) - \phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta))\right] \, \mathrm{d}G(\theta,\xi) \geq 0.$$

By equation (1), this is equivalent to showing that, for any $q \in \mathcal{Q}$,

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \theta \cdot \left[q^{\mathrm{SB}}(\theta) - q(\theta)\right]\left[\frac{c+\xi}{\theta} + \phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta))\right] \, \mathrm{d}G(\theta,\xi) \leq 0.$$

To analyze the integral on the left-hand side, we divide the interval $[\underline{\theta}, \overline{\theta}]$ into two sets:

$$\begin{cases}
\Theta_{\mathrm{bd}} &:= \left\{\theta \in [\underline{\theta}, \overline{\theta}] : \overline{q}^{\mathrm{FB}}(\theta) \leq 0 \text{ or } \overline{q}^{\mathrm{FB}}(\theta) \geq A\right\}, \\
\Theta_{\mathrm{int}} &:= \left\{\theta \in [\underline{\theta}, \overline{\theta}] : 0 < \overline{q}^{\mathrm{FB}}(\theta) < A\right\}.
\end{cases}$$

Now, we make the following observations:

- If $\theta \in \Theta_{\mathrm{bd}}$, then $q^{\mathrm{SB}}(\theta) = 0$ or A; hence we have

$$\phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta)) = -\frac{c+\xi}{\theta}.$$

This implies that the integral evaluates to zero on Θ_{bd}:

$$\int_{\theta\in\Theta_{\mathrm{bd}}}\int_{\underline{\xi}}^{\overline{\xi}} \theta\cdot\left[q^{\mathrm{SB}}(\theta)-q(\theta)\right]\left[\frac{c+\xi}{\theta}+\phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta))\right]\,\mathrm{d}G(\theta,\xi)=0 \qquad \forall\, q\in\mathcal{Q}. \tag{A1}$$

- If $\theta\in\Theta_{\mathrm{int}}$, then $q^{\mathrm{SB}}(\theta)=\overline{q}^{\mathrm{FB}}(\theta)\in(0,A)$. Since $\overline{q}^{\mathrm{FB}}$ is non-decreasing by construction, it follows that Θ_{int} must be an interval. Moreover, since v is stricly convex on $(0,A)$, it follows that $-v'\circ\overline{q}^{\mathrm{FB}}$ is a non-decreasing function with image in $[-v'(0),-v'(A)]$.

  Next, by comparing the expressions of q^{SB} in Theorem 2 and q_0 in Lemma 2, we deduce that $-v'\circ\overline{q}^{\mathrm{FB}}$ is the weighted least squares projection (with weights equal to consumption types θ) of $\theta\mapsto-(c+\xi)/\theta$ onto the set $\mathcal{Q}_{\mathrm{int}}$ of non-decreasing functions whose images are in $[-v'(0),-v'(A)]$ over the interval Θ_{int}:

$$\begin{aligned}
&-v'\circ\overline{q}^{\mathrm{FB}}\in\operatorname*{arg\,min}_{q\in\mathcal{Q}_{\mathrm{int}}}\int_{\theta\in\Theta_{\mathrm{int}}}\int_{\underline{\xi}}^{\overline{\xi}}\frac{\theta}{2}\left[\frac{c+\xi}{\theta}+q(\theta)\right]^2\,\mathrm{d}G(\theta,\xi)\\
\iff\;&\overline{q}^{\mathrm{FB}}\in\operatorname*{arg\,min}_{q\in\mathcal{Q}}\int_{\theta\in\Theta_{\mathrm{int}}}\int_{\underline{\xi}}^{\overline{\xi}}\frac{\theta}{2}\left[\frac{c+\xi}{\theta}-v'(q(\theta))\right]^2\,\mathrm{d}G(\theta,\xi).
\end{aligned}$$

  Thus $\overline{q}^{\mathrm{FB}}$ satisfies the variational inequality (cf. Theorem 1.4.2 of Balakrishnan, 1981)

$$\begin{aligned}
&\int_{\theta\in\Theta_{\mathrm{int}}}\int_{\underline{\xi}}^{\overline{\xi}}\theta\cdot\left[\overline{q}^{\mathrm{FB}}(\theta)-q(\theta)\right]\left[\frac{c+\xi}{\theta}-v'(\overline{q}^{\mathrm{FB}}(\theta))\right]\,\mathrm{d}G(\theta,\xi)\le 0\\
\iff\;&\int_{\theta\in\Theta_{\mathrm{int}}}\int_{\underline{\xi}}^{\overline{\xi}}\theta\cdot\left[q^{\mathrm{SB}}(\theta)-q(\theta)\right]\left[\frac{c+\xi}{\theta}+\phi^{(\theta,\xi)}(q^{\mathrm{SB}}(\theta))\right]\,\mathrm{d}G(\theta,\xi)\le 0\quad\forall\,q\in\mathcal{Q}.
\end{aligned}\tag{A2}$$

Finally, the desired inequality is obtained by adding equations (A1) and (A2). This completes the proof of Lemma 3.

A.4 Proof of Proposition 3

Given an individual demand curve $D(\cdot;\theta)$, observe that the deadweight loss due to that consumer consuming at a marginal price of $p(\theta)$ can be written as

$$\begin{aligned}
\mathrm{DWL}(\theta,\xi) &= \int_{p(\theta)}^{c+\xi} \left[D(p(\theta);\theta) - D(s;\theta)\right]\ \mathrm{d}s \\
&= \mathrm{CS}(c+\xi;\theta) - \mathrm{CS}(p(\theta);\theta) + \left[c+\xi-p(\theta)\right] D(p(\theta);\theta) \\
&= \Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, p(\theta)\right).
\end{aligned}$$

The expression for deadweight loss in Proposition 3 follows by summing over all consumers. Notice that this identity does not depend on functional form assumptions on individual demand curves.

A.5 Proof of Proposition 4

Let τ^* be the optimal linear tax. By Proposition 3, observe that $\tau^* = \mathcal{E}[\xi]$ and that the deadweight loss due to the optimal linear tax is $\mathcal{V}[\xi]$. Using the law of cosines for Bregman divergences, observe that the deadweight loss due to the optimal linear tax for a consumer with consumption type θ is

$$\begin{aligned}
\Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, c+\mathcal{E}[\xi]\right) = \Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, c+\mathbf{E}[\xi\,|\,\theta]\right) + \Delta^{\theta}_{\mathrm{CS}}\left(c+\mathbf{E}[\xi\,|\,\theta], c+\mathcal{E}[\xi]\right)& \\
+ \left(\xi - \mathbf{E}[\xi\,|\,\theta]\right)\left[D(c+\mathcal{E}[\xi];\theta) - D(c+\mathbf{E}[\xi\,|\,\theta];\theta)\right]&.
\end{aligned}$$

Taking expectation with respect to $\xi\,|\,\theta$ yields an alternative expression for $\mathcal{V}[\xi]$:

$$\begin{aligned}
\mathcal{V}[\xi] &= \underbrace{\mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, c+\mathcal{E}[\xi]\right)\right]}_{=\mathbf{E}[\mathrm{DWL}(\theta,\xi)]} \\
&= \mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, c+\mathbf{E}[\xi\,|\,\theta]\right)\right] + \mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}\left(c+\mathbf{E}[\xi\,|\,\theta], c+\mathcal{E}[\xi]\right)\right] \\
&= \mathbf{E}\left[\mathbf{E}\left[\Delta^{\theta}_{\mathrm{CS}}\left(c+\xi, c+\mathbf{E}[\xi\,|\,\theta]\right)\,|\,\theta\right]\right] + \mathcal{V}\left[\mathbf{E}[\xi\,|\,\theta]\right] \\
&= \mathbf{E}\left[\mathcal{V}[\xi\,|\,\theta]\right] + \mathcal{V}\left[\mathbf{E}[\xi\,|\,\theta]\right].
\end{aligned}$$

Here, the penultimate equality follows from the law of iterated expectations, and the definitions of $\mathcal{E}[\cdot]$, $\mathcal{V}[\cdot]$, and $\mathcal{V}[\cdot\,|\,\theta]$ are as given in the main text. When the second-best allocation function is strictly, the interpretations of $\mathbf{E}[\mathcal{V}[\xi\,|\,\theta]]$ and $\mathcal{V}[\mathbf{E}[\xi\,|\,\theta]]$ follow from Corollary 2, which implies that the second-best nonlinear tax must be $\tau^{\mathrm{SB}}(\theta) = \mathbf{E}[\xi\,|\,\theta]$.

A.6 Proof of Proposition 5

Let W be the distribution as defined in Theorem 2, and recall from Theorem 2 that

$$\overline{q}^{\mathrm{FB}}(\theta) := (v')^{-1}\left(\frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_0^s \frac{c+\mathbf{E}\left[\xi \,|\, \theta = W^{-1}(z)\right]}{W^{-1}(z)}\,\mathrm{d}z\right)\bigg|_{s=W(\theta)}\right).$$

Since $\overline{q}^{\mathrm{FB}}$ is non-decreasing, it is clear that a ban is optimal if and only if $\overline{q}^{\mathrm{FB}}(\overline{\theta}) \leq 0$. Now, notice that this condition is equivalent to

$$\begin{aligned}
&\overline{q}^{\mathrm{FB}}(\overline{\theta}) \leq 0 \\
\iff\;& \frac{\mathrm{d}}{\mathrm{d}s}\left(\mathrm{co}\int_0^s \frac{c+\mathbf{E}\left[\xi \,|\, \theta = W^{-1}(z)\right]}{W^{-1}(z)}\,\mathrm{d}z\right)\bigg|_{s=1} \geq v'(0) && \text{(since } v' \text{ is decreasing)} \\
\iff\;& \int_s^1 \left[\frac{c+\mathbf{E}\left[\xi \,|\, \theta = W^{-1}(s)\right]}{W^{-1}(z)} - v'(0)\right]\,\mathrm{d}z \geq 0 && \forall\, s \in [0,1] \\
\iff\;& \int_{\hat{\theta}}^{\overline{\theta}} \int_{\underline{\xi}}^{\overline{\xi}} \left(c+\xi-\theta v'(0)\right)\,\mathrm{d}G(\theta,\xi) \geq 0 && \forall\, \hat{\theta} \in [\underline{\theta}, \overline{\theta}].
\end{aligned}$$

This proves part (i) of Proposition 5; part (ii) of Proposition 5 can be proved similarly by using the fact that a mandate is optimal if and only if $\overline{q}^{\mathrm{FB}}(\underline{\theta}) \geq 0$.

A.7 Proof of Proposition 6

Define

$$\Psi(s) := \int_0^s \frac{c+\mathbf{E}\left[\xi \,|\, \theta = W^{-1}(z)\right]}{W^{-1}(z)}\,\mathrm{d}s.$$

Observe that a quantity floor is optimal if and only if there exists $s \in (0,1]$ such that $\Psi(s) \geq s \cdot \Psi'(0)$. This condition is equivalent to

$$\begin{aligned}
&\exists\, \theta^* \in (\underline{\theta}, \overline{\theta}] : && \int_{\underline{\theta}}^{\theta^*} \left[\frac{c+\mathbf{E}\left[\xi \,|\, \theta = s\right]}{s} - \frac{c+\mathbf{E}\left[\xi \,|\, \theta = \underline{\theta}\right]}{\underline{\theta}}\right]\,\mathrm{d}W(s) \geq 0 \\
\iff\;&\exists\, \theta^* \in (\underline{\theta}, \overline{\theta}] : && \int_{\underline{\theta}}^{\theta^*} \int_{\underline{\xi}}^{\overline{\xi}} \left[c+\xi-\theta \cdot \frac{c+\mathbf{E}\left[\xi \,|\, \theta = \underline{\theta}\right]}{\underline{\theta}}\right]\,\mathrm{d}G(\theta,\xi) \geq 0 \\
\iff\;&\exists\, \theta^* \in (\underline{\theta}, \overline{\theta}] : && \frac{c+\mathbf{E}\left[\xi \,|\, \theta \leq \theta^*\right]}{c+\mathbf{E}\left[\xi \,|\, \theta = \underline{\theta}\right]} \geq \mathbf{E}\left[\frac{\theta}{\underline{\theta}} \,|\, \theta \leq \theta^*\right].
\end{aligned}$$

This proves part (i) of Proposition 6; part (ii) of Proposition 6 can be proved similarly by using the fact that a mandate is optimal if and only if $\Psi(1) - \Psi(s) \leq (1-s) \cdot \Psi'(1)$.

A.8 Proof of Corollary 3

Let Ψ be as defined in the proof of Proposition 6. To prove part (i) of Corollary 3, it suffices to show that

$$\chi'(\underline{\theta}) > 0 \implies \Psi(s) \geq s \cdot \Psi'(0).$$

To this end, observe that

$$\Psi'(s) = \frac{c + \mathbf{E}\left[\xi \mid \theta = W^{-1}(s)\right]}{W^{-1}(s)}.$$

Since W is increasing, it follows that $\chi'(\underline{\theta}) \geq 0 \implies \Psi'(0)$ is increasing. Therefore $\Psi(s)$ is strictly convex in a neighborhood of $s = 0$, from which it must be that $\Psi(s) \geq s \cdot \Psi'(0)$. Part (ii) of Corollary 3 follows from a similar argument.