Correct Me if You Can - Optimal Non-Linear Taxation of Internalities*

Andreas Gerster† and Michael Kramm‡

September 14, 2018

Abstract

Consumers' private valuations of a product characteristic can be systematically biased, leading to welfare losses from what are referred to as "internalities". Previous evidence has shown that these kinds of behavioral failures affect consumer choices in many settings such as energy efficiency investments concerning vehicles or housing, for instance, or the consumption of sugar-sweetened beverages. In this paper, we characterize the optimal non-linear tax (or subsidy) for correcting behaviorally biased consumers. Using representative examples we show that it is welfare-enhancing compared to no taxation and linear taxation. We examine how the amount of welfare improvement delivered depends on informational restrictions affecting the mechanism designer. Furthermore, we briefly contrast our result with subsidy schemes for fostering energy efficiency in the German housing stock and show that the optimal tax is essentially antipodal to current practice.

Keywords: optimal commodity taxation, nonlinear taxation, behavioral economics, public economics, internalities, environmental economics

JEL codes: H21, D82, D03, Q58

*We would like to express our gratitude for the discussions with (in alphabetical order) Jason Abaluck, Jörg Peter Lingens, and Lars Metzger.

†Corresponding Author, RWI - Leibniz-Institut für Wirtschaftsforschung, Hohenzollernstraße 1-3, 45128 Essen, Germany, Tel.: +49 (201) 8149-521, mail: andreas.gerster@rwi-essen.de

‡Technical University of Dortmund and Ruhr Graduate School in Economics, mail: michael.kramm@tu-dortmund.de

1 Introduction

Consumer sovereignty has long been undebated among economic scholars: the primacy of rationality implies that consumers have well-behaved preferences, and that of Bayes rationality that new information is processed in an optimal way. Over the past decades, researchers discuss the practical and theoretical implications of violations of these fundamental assumptions. Behavioral economics suggests that if agents do not act in their best interest, then there is room for intervention by a benevolent policy maker. The growing literature on behavioral mechanism design and behavioral public economics examines this issue.

Misoptimizing consumers inflict a so-called “internality” upon themselves, which provides a justification for corrective taxation beyond the classical case of externalities. An important feature of internalities is that they are inherently heterogeneous as they depend on an individual’s bias, in contrast to the social cost of global externalities, for example, which do not vary by agent. As a consequence, targeting corrective taxes towards behaviorally biased consumers to improve welfare is particularly important in the case of internalities. Furthermore, in contrast to externalities, individuals’ choices reflect their unobserved bias, which can be exploited for targeting.

In this paper, we explore the potential of non-linear commodity taxation to target behaviorally biased consumers and to increase social welfare. Following Farhi and Gabaix (2017) and Mullainathan et al. (2012), we employ a general model of biases that encompasses a broad class of behavioral failures driving a wedge between “experienced” and “decision utility” (Kahneman et al., 1997), such as limited attention, biased beliefs and present bias. Based on this generic specification of a behavioral bias, we derive the optimal non-linear commodity tax using a mechanism design approach. Our approach allows for commodity taxation in a broad sense that encompasses non-linear taxation of quantities, but also of qualities, which is important as internalities often distort choices between product varieties. For example, consumers fail to take product attributes into account that are less salient than the purchase price, such as co-pays of a life insurance policy and the operating cost of a durable (Abaluck and Gruber, 2011; Allcott and Taubinsky, 2015).

Starting with Pigou (1920), corrective taxation to overcome market failures from externalities has been studied extensively. Recently, a growing literature has studied optimal taxation for behaviorally biased consumers (O’Donoghue and Rabin, 2006; Allcott et al., 2015; Farhi and Gabaix, 2017), yet the potential to target behaviorally biased consumers through non-linear taxation has not been investigated so far. Traditionally, public tax schemes have targeted consumers through “tagging” (Akerlof, 1978), i.e., a conditioning of taxes on observable characteristics. While appealing in theory, it is difficult to avoid strategic behavior and to find immutable characteristics that are socially acceptable. In contrast, a large literature in industrial organization has investigated how firms can use price discrimination to maximize profits (Mussa and Rosen, 1978). We combine these strands of the literature and explore the potential of non-linear tax schemes as a means to correct behaviorally biased consumers and to increase social welfare.

We show that non-linear commodity taxes increase welfare beyond the optimal constant per-unit tax. In particular, when perceived valuations and biases are positively correlated, marginal taxes should be lower for types with a low perceived valuation, and higher for high types. The intuition for this result is straigthforward and parallels the findings in the price discrimination literature: when confronted with a continuous range of quality, consumers partly reveal their type through their position on the demand curve, which can be exploited for targeting. We show how the scope of welfare improvement due to a corrective taxation changes with the information about aggregate consumer misoptimization available to the mechanism designer. Furthermore, we discuss empirical and theoretical findings on the correlation between perceived valuations and biases and contrast our optimal non-linear tax scheme with a behaviorally motivated policy in Germany.

The paper is structured as follows. In the next section we discuss the related literature and in Section 3 we introduce the model setup. Section 4 then contains the analytical characterization of the optimal tax scheme, while Section 5 discusses illustrative examples. Section 6 discusses our findings and concludes. All proofs and the figures for the numerical simulations discussed in Section 5 can be found in the Appendices.

2 Related Literature

Our model is related to all three main branches of the research on optimal taxation.[1] First, since we study the taxation of a consumption good such as an electricity-using durable, our analysis examines commodity taxation. This field of research was originated by Ramsey (1927) and its main issue is to study how to distribute linear taxation across different commodities in order to obtain a certain government budget with the least harming consumption distortions.

Second, the study of corrective taxation in order to combat market failures due to externalities was initiated by Pigou (1920). We study corrective taxation with internalities. The role of heterogeneity among consumers in the valuation of a public good and the associated externalities agents might inflict on others by free-riding was recognized for instance in the analysis of optimally individualized Lindahl prices, see Lindahl

[1] An excellent overview on the research on optimal taxation is given in Salanié (2011).

(1919). This approach did not account for the informational asymmetries, which exist among a social planner designing the tax and the consumers who hold private information on their valuations or abilities.

The third main field of optimal taxation our work relates to - especially on a methodological basis - is that of non-linear income taxation. Mirrlees (1971) was the first to provide a method to cope with the above mentioned informational asymmetries between the designer and consumers. He explicitly accounted for the consumers' strategic responses to taxes and is one of the founding fathers of mechanism design due to examining incentive compatible mechnisms. The literature on optimal income taxation focusses on the equity-efficiency trade-off. It analyzes how to redistribute income with imperfect knowledge about the agents labor skills by taxing income non-linearly.[2] Two major continuations in optimal non-linear taxation are Diamond (1998) and Saez (2001). Nonlinear commodity taxation has been studied, for instance, by Mirrlees (1976) and Cremer and Gahvari (1998), and nonlinear taxation with externalities has been studied by Cremer et al. (1998), for instance.

The pathbreaking methodological advances of Mirrlees (1971) have also had an immense impact outside the realm of public economics. The theory of non-linear pricing contrasts optimal non-linear taxation in that the designer's objective no longer is the maximization of social welfare, but rather the maximization of profit.[3] However, the fundamental commonality is that the designer has incomplete information about the preferences of the individual consumer and has to take this into account when designing the mechanism.

Behavioral economics and also behavioral game theory have become a growing field of research, see Kőszegi (2014) and Camerer (2003). The same holds true for behavioral public economics, see Oliver (2013), Sunstein (2016), Congdon et al. (2011), Rees-Jones and Taubinsky (2017), and Chetty (2015). Mullainathan et al. (2012) develope a general model for behavioral public policy. Research on behavioral public policy with a focus on environmental economics can be found in Allcott (2016) and Allcott et al. (2017). Behavioral income taxation is studied in Lockwood and Taubinsky (2016) and Gerritsen (2016). A behavioral approach to commodity taxation with internalities and externalities is pursued in Allcott et al. (2014), Allcott et al. (2015), and Lockwood and Taubinsky (2017). The work most closely related to ours is Allcott and Taubinsky (2015), as we discuss below (at the end of Section 3 and in Section 4 after Proposition 1).

[2]Generally, incentive compatibility and externalities are closely connected. In the second-price auction, for instance, an incentive compatible auction mechanism, the winner has to pay the second-highest bid, which is a measure of how much harm the winner of the auction inflicts upon society, i.e., it takes into account the externalities of the highest bid.

[3]The theory of non-linear pricing is surveyed in Wilson (1997).

3 Model Setup

The main goal of this section is to establish a model that allows us to analyze how the social planner (*she*), i.e. the mechanism designer, implements a welfare maximizing tax scheme - or equivalently a subsidy scheme - in an economy with behaviorally biased consumers (all *he*).

We model the interaction between the mechanism designer and the consumers as a dynamic Bayesian game with two stages. In period one, the designer commits to a (possibly) non-linear tax regime $t : X \to \mathbb{R}$, where $X \subseteq [0, \infty)$ is the consumer's choice set, which is identical to each of them. In period two, the consumers then simultaneously choose their consumption $x \in X$. The game is solved by backward induction. We begin by presenting the characteristics of the consumer side, and then describe the problem of the mechanism designer.

3.1 Consumers

There is a unit mass of consumers. We refrain from indexing a specific consumer $i \in [0, 1]$ whenever possible for the ease of notation. A consumer's choice variable is given by $x \in X \subseteq [0, \infty)$. A consumer's "true" per-unit valuation of good x is captured by the random variable v, which is distributed according to the cumulative distribution function F with support $supp(F) := [\underline{v}, \overline{v}] \subseteq (-\infty, \infty)$, $-\infty < E[v] < \infty$, and density function f. The bias b, reflecting a misperception of true valuation of the consumption choice x, is distributed according to the cumulative distribution function G with support $supp(G) := \left[\underline{b}, \overline{b}\right] \subseteq (-\infty, \infty)$, $-\infty < E[b] < \infty$, and with the density g. We assume that v and b are i.i.d. across consumers and each other.

A consumer's perceived per-unit valuation of good x is given by $\hat{v} : [\underline{v}, \overline{v}] \times \left[\underline{b}, \overline{b}\right] \to \mathbb{R}$, $(v, b) \mapsto \hat{v}(v, b)$, which depends on the true valuation v and the bias b. We assume that the perceived valuation increases in the true valuation and in the bias. Furthermore, we specify the bias such that $\hat{v}(v, 0) = v$, i.e. a consumer with bias $b = 0$ is unbiased, while $b < 0$ ($b > 0$) imply underestimation (overestimation) of the value of consumption. The perceived valuation $\hat{v}$ is distributed according to the cumulative distribution function P, which is induced by the distributions of v and b. The density is given by p. The support $supp(P) := \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ is determined by the support of the distributions F and G. The setup immediately implies that the perceived valuation $\hat{v}$ is i.i.d. across consumers. We present examples for the dependence of the functional form of P on the functional form of F and G in Section 5. We assume that consumers only observe their own perceived valuation $\hat{v}(v, b)$ and do not in any form update their belief about their own true valuation.

Let $z \in \mathbb{R}$ denote the numeraire good, i.e., the money a consumer spends for the consumption of other goods. The consumer's money available for the consumption of goods other than x given choice x and tax scheme t is given by $z(x,t) := w(x) - t(x)$ with $w : X \to \mathbb{R}$ concave. The consumer's objective is given by his decision utility $u^d : X \times \mathbb{R} \times [\underline{\hat{v}}, \overline{\hat{v}}] \to \mathbb{R}$ with $(x, z, \hat{v}) \mapsto u^d(x, z, \hat{v})$. The consumer's experienced utility is given by $u^e : X \times \mathbb{R} \times [\underline{v}, \overline{v}] \to \mathbb{R}$ with $(x, z, v) \mapsto u^e(x, z, v)$. Experienced utility increases in the true valuation and decision utility increases in the perceived valuation. Experienced and decision utility increase in z. The individual choice of x is observable.

In order to obtain explicit solutions, we follow a common routine in the literature and assume quasilinear utility with an increasing and convex cost function $c : X \to \mathbb{R}$. We assume that good x is produced on competitive markets so that the cost function corresponds to the price of consuming x. Let $m \in \mathbb{R}_{++}$ denote an initial endowment with the numeraire. Furthermore, we assume an additive bias $\hat{v} := v + b$. This implies that the decision utility can be written as

$$u^d(x, \hat{v}) = m + \hat{v}x - t(x) - c(x),$$

and the experienced utility as

$$u^e(x, v) = m + vx - t(x) - c(x).$$

The quasi-linearity of $u^d(\cdot)$ – and $u^e(\cdot)$ analogously – becomes obvious if we write it as $u^d(x, z, \hat{v}) = z - c(x)$, which is linear in z and non-linear in x, and assume that the budget constraint is given by $z = m + \hat{v} \cdot x - t(x)$. Note that $u^e(x, z, v) = u^d(x, z, \hat{v}) - bx$.

3.2 Mechanism Designer

The designer's objective is to elaborate a tax scheme $t : X \to \mathbb{R}$, based on information about the distributions P, F, and G. She observes the consumers' choices x, but not any realization of the random variables v, b, and $\hat{v}$. The designer's objective function consists of the increasing and concave social welfare function $W : \mathbb{R}^{[0,1]} \to \mathbb{R}$ with $\{u_i^e\}_{i\in[0,1]} \mapsto W\left(\{u_i^e\}_{i\in[0,1]}\right)$. For simplicity, we assume $\{u_i^e\}_{i\in[0,1]} \mapsto W\left(\{u_i^e\}_{i\in[0,1]}\right) := \int_0^1 \alpha(v_i) u_i^e di$ with $\alpha : [\underline{v}, \overline{v}] \to \mathbb{R}_+$. This implies that individual i's experienced utility is weighted with $\alpha(v_i)$, which does not take into account i's identity, but rather his true valuation.

Letting $\mathbb{T} := \{f | f : X \to \mathbb{R}\}$ denote the function space containing all functions between domain X and codomain $\mathbb{R}$, the objective of the designer is thus given by

$$\max_{t\in\mathbb{T}} \int_0^1 \left(\int_v \alpha(v_i) \cdot u_i^e \left(x_i, z_i(x_i, t), v_i\right) f(v) dv \right) di + \int_0^1 t(x_i) di.$$

3.3 Discussion of the Model Setup

The perceptional bias manifests itself in the fact that consumers experience a heterogeneous internality upon consuming or using the good under consideration, e.g., an energy-using durable or a sugar-sweetened beverage. Although externalities and internalities both imply that a social planner can increase social welfare, for instance, via taxation, targeting through non-linear taxes is only feasible for internalities. Welfare losses from externalities arise from the fact that they are not reflected in the individuals' decisions. In contrast, internalities reduce welfare by (wrongly) influencing decision utility. As a consequence, observing consumption choices is only informative about internalities, but not externalities.

To fix ideas, assume that a consumer chooses one of several varieties of a horizontally differentiated product, such as a electricity-using durable of varying energy efficiency levels. In this example, x can be interpreted as the energy efficiency of the durable, measured relative to the worst variety on the market, and v corresponds to the actual reduction in operating cost through better energy efficiency, given actual usage and prices. The random variable b can be interpreted as a misperception of the value of energy efficiency at the time of purchase, i.e., a wedge between perceived valuation $\hat{v}$ and the "true" valuation v of energy efficiency. We allow the bias to have any (finite) magnitude in expectation, but it is important to note that empirical applications have typically found that consumers are biased downwards, which means that they undervalue certain attributes, for instance due to limited attention (Allcott et al., 2015). An additive specification of the bias is common in the literature, see Farhi and Gabaix (2017) and Mullainathan et al. (2012).[4]

The departure from Bayesian rationality that is inherent in our model, i.e., the fact that consumers perceive only $\hat{v}$ and do not make inferences on their bias, can be explained by one of the following behavioral failures that found extensive support in the literature. First, present biases can induce consumers to undervalue products that pay out only in the future, such as healthy foods and energy efficient durables (Laibson, 1997; Loewenstein and Prelec, 1992; O'Donoghue and Rabin, 1999, 2006). Second, if consumers pay only limited attention to certain product attributes, decision utility does not match experienced utility. Inattention has been documented in a vast variety of settings including energy efficiency (Allcott and Taubinsky, 2015) and health care choices (Abaluck and Gruber, 2011). Third, biased beliefs can equally drive a wedge between experienced and decision utility. For example, the literature has found evidence for biased beliefs with regard to energy efficiency, the calory content of nutrition, and schooling returns (Attari et al., 2010; Bollinger et al., 2011; Jensen, 2010).

[4]The main arguments remain the same for a multiplicative specification of the bias.

The assumption of quasilinear utility implies that the model abstracts from income effects. These are not the focus of this paper, but are extensively discussed in the literature on optimal income taxation. Note also that the functional form of the consumer's objective with respect to x is endogenous to the model. The related issue of the existence of interior maximizers of consumer utility is discussed in Appendix 7.A.

To isolate the corrective nature of taxation in our setting, we assume that the designer can levy lump-sum taxes. As a result, the marginal utility of public funds is unity and the designer has no incentive to distort prices to raise public funds. In addition, owing to our assumption that lump-sum taxation is feasible, we do not need to model a government resource constraint.

As it is common in the mechanism design literature, using the standard solution techniques requires that the hazard rate of the type distribution is increasing. In our setting, the consumers' (perceived) types are given by $\hat{v}$; its distribution is determined by a convolution of the distributions v and b, see Bertsekas and Tsitsiklis (2008, pp. 213). While this is a deviation from standard assumptions in the mechanism design literature, its implications are only minor. In particular, Barlow et al. (1963) show that the set of distributions with an increasing hazard rate is closed under convolution. This implies that if the distributions of v and b have an increasing hazard rate, so does the distribution of $\hat{v}$.[5]

As we will see later on when solving the model, a central distinction to the standard mechanism design problems is the fact that the *experienced utility* is evaluated at x^*, the solution to the maximization problem involving the (possibly biased) *decision utility.* In Section 4, we analyze how the designer can possibly correct the internality to increase social welfare using non-linear taxation. Inspired by Allcott and Taubinsky (2015), who examine corrective *linear* commodity taxes in a framework with internalities and a *binary* decision, we investigate the novel question of how *non-linear* tax instruments affect the welfare of heterogeneous consumers facing a *continuous* choice set. In contrast to the existing literature on behavioral commodity taxation, this automatically introduces the issue of incentive compatibility into our setup, as consumers have an incentive to misreport their type.

4 The Optimal Tax Scheme with Quasilinear Utility

In this section, we apply the concept of a perfect Bayesian Nash equilbirium to derive the optimal tax scheme. We proceed in two steps. First, we solve for the optimal behavior of behaviorally biased consumers that maximizes decision utility (but not

[5] Furthermore, alternative solution techniques are available if the increasing hazard rate property is not satisfied, such as the Myerson-ironing approach.

necessarily experienced utility). Second, we derive the optimal tax schedule and discuss its properties.

The behavior of the biased consumer is captured by

$$x^*(\hat{v}, t) := \arg\max_x u^d(x, z(x,t), \hat{v}),$$

and that of the unbiased consumer by

$$x^u(v, t) := \arg\max_x u^e(x, z(x,t), v).$$

To simplify notation, we sometimes write x^* instead of $x^*(\hat{v}, t)$ and z^* instead of $z(x^*, t)$, and we proceed in the same manner with x^u and z^u. The model will generically imply a loss in consumer surplus $CS := \int_i u_i^e(x_i, z_i, v_i)di$ compared to a setup with non-biased consumers, as $u^e(x^*, z^*, v) \leq u^e(x^u, z^u, v)$. Due to Bayesian irrationality, even in the case of $E[b] = 0$, consumers misoptimize in expectation and a loss in consumer surplus occurs, as we discuss in Section 5.

Börgers (2015) paraphrases the Revelation Principle as: "If an allocation can be implemented through some mechanism, then it can also be implemented through a direct truthful mechanism where the consumer reveals his information about his type". In contrast to a conventional mechanism design approach, the consumers do not know their true valuation in the setup at hand, but decide exclusively on the basis of perceived valuations $\hat{v}$. That is, in a straightforward extension of the Revelation Principle to our setup, the type of a consumer is given by his perceived valuation $\hat{v}$, rather than by (v, b). The designer confines herself to designing a direct mechanism $(\xi, \tau) : \left[\underline{\hat{v}}, \overline{\hat{v}}\right] \to X \times \mathbb{R}$ under truthtelling to implement the desired outcome. Based on a consumer's strategical report $\tilde{v}$, the direct mechanism then assigns the consumed quantity, $\xi(\tilde{v}) \in X$, and the amount of taxes to be paid, $\tau(\tilde{v}) \in \mathbb{R}$.

Under the direct mechanism, the decision utility for report $\tilde{v}$ given the perceived valuation $\hat{v}$ is

$$u^d(\xi(\tilde{v}), z(\tilde{v})|\hat{v}) = m + \hat{v} \cdot \xi(\tilde{v}) - \tau(\tilde{v}) - c(\xi(\tilde{v})).$$

Since the consumer may strategically misreport his perceived valuation, truthtelling can be induced by the designer by implementing an incentive compatible mechanism (in dominant strategies). This implies that the tax scheme must satisfy

$$u^d(\xi(\hat{v}), z(\hat{v})|\hat{v}) \geq u^d(\xi(\tilde{v}), z(\tilde{v})|\hat{v}) \;\forall \hat{v}, \tilde{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]. \tag{IC}$$

As in the standard mechanism design setup, ξ must be increasing in $\hat{v}$ for direct mechanisms to be valid. A standard sufficient condition for ξ to increase is that the hazard rate of P increases.

Optimal strategic reporting of a consumer implies that the solution v^* to the problem $\max_{\tilde{v}} u^d(\xi(\tilde{v}), z(\tilde{v})|\hat{v})$ has to satisfy

$$\hat{v}\xi'(v^*) - \tau'(v^*) - \xi'(v^*)c'(\xi(v^*)) \stackrel{!}{=} 0. \tag{1}$$

Incentive compatibility implies $v^* = \hat{v}$ and thus equilibrium decision utility in an incentive-compatible direct mechanism is given by $\hat{u}^d(\hat{v}) := u^d(\xi(\hat{v}), z(\hat{v})|\hat{v})$, while equilibrium experienced utility is given by $\hat{u}^e(\hat{v}) := u^e(\xi(\hat{v}), z(\hat{v})|v) = \hat{u}^d(\hat{v}) + b\xi(\hat{v})$. Put differently, incentive compatibility implies that for all $\hat{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ it has to hold that

$$\tfrac{\partial \hat{u}^d(\hat{v})}{\partial \hat{v}} = \xi(\hat{v}) + \hat{v}\xi'(\hat{v}) - \tau'(\hat{v}) - \xi'(\hat{v})c'(\xi(\hat{v})) \stackrel{(1)}{=} \xi(\hat{v}). \tag{2}$$

We assume that the outside option of any consumer is such that he always prefers to participate in the mechanism. Therefore, the participation constraints are fulfilled for any (perceived) consumer type. As usual, when incentive compatibility is satisfied for each perceived type, then it suffices that the participation constraint of the lowest type is satisfied. In particular we can specify the model such that $\hat{u}^d\left(\underline{\hat{v}}\right) = \underline{u} > 0$ and $\hat{u}^d\left(\overline{\hat{v}}\right) \geq \underline{u}$.

The designer solves a dynamic optimization problem which can be analyzed using the optimal control approach.[6] Note that determining the equilibrium values of $\xi(\hat{v})$ and $u^d(\hat{v})$ for all $\hat{v}$ pins down the equilibrium value of $\tau(\hat{v})$ for all $\hat{v}$. Hence, the mechanism design problem of the designer is given by

$$\max_{\xi \in \mathbb{X}} \int_0^1 \left(\int_v \left[\int_b \alpha(v_i) \cdot \hat{u}_i^e\left(\hat{v}_i\right) dG(b) \right] dF(v) \right) di + \int_0^1 \int_v \int_b \tau\left(\hat{v}_i\right) dG(b) dF(v) di, \tag{3}$$

subject to the condition from Equation (2), and where $\mathbb{X} := \{f | f : \left[\underline{\hat{v}}, \overline{\hat{v}}\right] \to X\}$ is the function space containing all functions between domain $\left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ and codomain X. The boundary conditions of the problem are given by $\hat{u}^d\left(\underline{\hat{v}}\right) = \underline{u}$ and $\hat{u}^d\left(\overline{\hat{v}}\right) \geq \underline{u}$. The control variable is ξ and the law of motion of the state variable $\hat{u}^d$ is given by incentive compatibility and optimal strategic reporting. Using the definition of decision utility

[6] For details on what follows consider Sydsaeter et al. (2008, papers 9-10).

to replace the tax and rewriting equilibrium experienced utility in terms of equilibrium decision utility, the Hamiltonian for the above problem for all $\hat{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ is given by

$$H\left(\hat{v}, \xi, \hat{u}^d\right) = \left[E[\alpha(v)|\hat{v}] \cdot \underbrace{\left(\hat{u}^d(\hat{v}) - E[b|\hat{v}]\xi(\hat{v})\right)}_{=\hat{u}^e(\hat{v})} + \underbrace{\left(m + \hat{v}\xi(\hat{v}) - \hat{u}^d(\hat{v}) - c(\xi(\hat{v}))\right)}_{=\tau(\hat{v})}\right] p(\hat{v}) + \mu(\hat{v})\xi(\hat{v}).$$

Following the standard solution procedure for mechanism design problems, we employ Pontryagin's Maximum Principle, which yields the following necessary conditions for the optimal tax.[7]

$$\text{FOC on control: } \frac{\partial H}{\partial \xi} = \left[-E[b|\hat{v}] \cdot E[\alpha(v)|\hat{v}] + \hat{v} - c'(\cdot)\right] p(\hat{v}) + \mu\left(\hat{v}\right) \stackrel{!}{=} 0, \quad (\text{FOC}_\text{x})$$

$$\text{FOC on state: } \frac{\partial H}{\partial \hat{u}^d} = \left[E[\alpha(v)|\hat{v}] - 1\right] p(\hat{v}) \stackrel{!}{=} -\mu'\left(\hat{v}\right), \quad (\text{FOC}_\text{u})$$

$$\text{transversality cond.: } \mu\left(\underline{\hat{v}}\right) \cdot \hat{u}^d\left(\underline{\hat{v}}\right) = \mu\left(\overline{\hat{v}}\right) \cdot \hat{u}^d\left(\overline{\hat{v}}\right) = 0. \quad (\text{TVC})$$

We obtain the characterization of the optimal non-linear tax scheme in our main result:

Proposition 1. *The optimal non-linear commodity tax in the model with internalities and quasilinear utility is implicitly given by*

$$t'(x) = \frac{\int_{\hat{v}_x}^{\overline{\hat{v}}} \left(1 - E[\alpha(v)|m]\right) p(m) dm}{p(\hat{v}_x)} + E[b|\hat{v}_x] \cdot E[\alpha(v)|\hat{v}_x] \quad \forall x \in X, \tag{4}$$

where $\hat{v}_x$ is the report to be sent by a consumer to obtain the allocation x under the optimal tax scheme.

Proof. See Appendix 7.A □

Note that throughout we have assumed that v and b are independent. However, ceteris paribus, analogous results hold for cases in which true valuation v and bias b are correlated. For instance in the problem of the mechnism designer stated in Equation (3), we would need to use a conditional distribution of b depending on v, and in the optimal tax formula given by Equation (4) the conditional expectations and the induced density p would have to be modified accordingly. Otherwise, the arguments remain the same. We highlight this insight in the following corollary.

Corollary 1. *A result analogous to Proposition 1 holds for cases in which true valuation v and bias b are correlated.*

[7] Sufficiency is given if in addition the control region is convex and the Hamiltonian is concave in $(\xi, \hat{u}^d)$ for every $\hat{v}$, see Sydsaeter et al. (2008, page 315). This is satisfied in the setup at hand.

The results apply to any distribution of perceived valuations, which satisfies the increasing hazard rate property. In particular, this is the case for the convolution of two uniformly distributed independent random variables v and b, which yields a trapezoid distribution, as we discuss below in Section 5. Additionally, this is also the case for the convolution of two normally distributed independent random variables v and b, which is itself a normal distribution.

To convey the intuition of the optimal marginal tax we first examine the second summand of Equation (4). This term embodies the behavioral aspect of our model and does not appear in the standard literature on non-linear taxation. It corrects for the expected bias of a consumer conditional on the report about his perceived valuation. It may be negative as well as positive depending on the expected conditional bias. The designer uses her potential to help the consumers in correcting their Bayes irrationality with the optimal tax scheme - "she corrects them, if she can".

The optimal tax formular in the binary choice model with linear taxation of Allcott and Taubinsky (2015) does not allow for "price discrimination" among consumers. In contrast to this, in our model, since there is a correlation between bias and report the consumer's strategic report about his perceived valuation reveals information about his bias and the designer should exploit this by using Bayesian updating. Details on this updating process can be found in Section 5.

If the designer has a utilitarian social welfare function with equal weights for each consumer independent of his valuation, we can normalize these weights to one without loss of generality and $E[\alpha(v_i)|\hat{v}_x] = 1$ for all $\hat{v}_x \in [\underline{\hat{v}}, \overline{\hat{v}}]$. In this case, the marginal tax rate given by Equation (4) just equals the expected bias conditional on the report, $t'(x) = E[b|\hat{v}_x]$. In Appendix 7.B, we show that when the designer cannot discriminate among the different consumers and has to resort to linear taxation, the optimal tax is given by $t^* = E[b]$, that is, the marginal tax rate equals the *unconditional* expected bias. Furthermore, as indicated above, Allcott and Taubinsky (2015) have shown that in a binary investment setting the optimal tax is equal to the average bias of the consumers who are indifferent between both goods at market prices. Again, our result on the non-linear tax differs by its dependence on consumers reports $\hat{v}$ rather than a fixed market price. In fact, it is exactly the variation in reports that a mechanism designer can exploit with a non-linear tax scheme to improve upon a constant per-unit tax.

Next, we contrast our optimal non-linear tax for behaviorally biased consumers with the famous ABC formula from the theory of optimal non-linear income taxation, derived by Diamond (1998, p. 86, Equation (10)). The ABC formula contains three factors: efficiency considerations (A), redistribution issues (B), and the dependence of the incentive compatibility constraint on the density functions via the hazard rate (C). To start with, part A of the ABC formula is not present in our model, which reflects that our model abstracts from income effects. Yet, redistribution issues (B) and the density of $\hat{v}$ (C) are contained in the first summand of Equation (4), although in modified form. The intuition of this summand is as follows. When the designer changes the marginal tax at, say, x, she extracts money from all consumers with $\hat{v} \geq \hat{v}_x$. The change of her objective is captured by the term $\int_{\underline{\hat{v}}}^{\overline{\hat{v}}_x} (1 - E[\alpha(v_i)|m])\, p(m)dm$, since the marginal value of an additional unit of tax money is one and welfare decreases by $E[\alpha(v_i)|\hat{v}_x]$ for a consumer with type $\hat{v}_x$. Intuitively, if the average welfare weights $\alpha(v)$ for these consumers exceed unity (and thus the marginal value of the tax income to the designer), the designer's objective function decreases and she reduces the tax. This term is weighted more strongly in the optimal tax formula if the density of the type $\hat{v}_x$ is low, i.e. if only few consumers are marginal to the tax change at x and thus have incentives to change their behavior. The second summand in our optimal tax formula captures the novel aspect of corrective taxation in our model and does not appear in the standard ABC formula. Another notable difference to the the standard models without behavioral consumers is that our result allows for a non-zero marginal tax rate at the top of the type distribution.

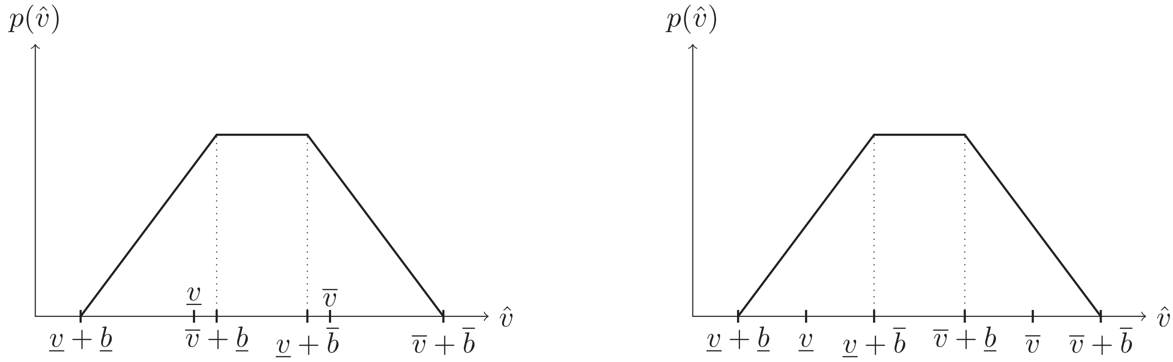
5 Illustrative Examples: Updating, Optimal Tax, and Welfare

We now make distributional assumptions and explore the inferences that a designer can make about consumer types. Furthermore, we derive the optimal non-linear tax schedules for these examples and discuss their properties. For the ease of exposition, we assume that the designer has a utilitarian social welfare function with equal weights normalized to one for each consumer, so that, according to Proposition 1, the optimal marginal tax rate just equals the expected bias conditional on the report, i.e. $t'(x) = E[b|\hat{v}_x]$.

5.1 Density $p(\hat{v})$ of the Perceived Valuation

To explore how P is determined by the two random variables v and b, we assume that they are independently and uniformly distributed: $b \sim U\left[\underline{b}, \overline{b}\right]$ and $v \sim U[\underline{v}, \overline{v}]$. Furthermore, we assume that $\overline{b} < \underline{v}$, i.e., the realization of the true valuation is always larger than that of the bias. Note that uniform distributions exhibit the increasing-hazard-rate property and by the above mentioned result of Barlow et al. (1963) this property is inherited by their convolution.

There are two distinct general cases: either the variance, and thus the entropy,[8] of the bias is smaller than that of the true valuation, implied by $\overline{v} - \underline{v} \geq \overline{b} - \underline{b}$, or vice versa, $\overline{v} - \underline{v} < \overline{b} - \underline{b}$. Figure 1 visualizes that in both cases the induced density p follows a trapezoid distribution (in the case of $\overline{v} - \underline{v} = \overline{b} - \underline{b}$ it follows a triangular distribution, which is a special trapezoid distribution). The trapezoid distribution consists of three ranges: a triangular lower part, a rectangular middle part, and a triangular upper part. If the variance of the bias exceeds the variance of the valuation, the middle range spans over the interval $\left[\overline{v} + \underline{b}, \underline{v} + \overline{b}\right]$, and $\left[\underline{v} + \overline{b}, \overline{v} + \underline{b}\right]$ otherwise. This has important implications for the updating process the designer can exploit, as we discuss below.



(a) Variance of valuation smaller than variance of bias: $\overline{v} - \underline{v} < \overline{b} - \underline{b}$:

(b) Variance of valuation greater or equal than variance of bias: $\overline{v} - \underline{v} \geq \overline{b} - \underline{b}$

Figure 1: Density p derived from the convolution of f and g when both, v and b, are i.i.d and uniformly distributed.

5.2 Conditional Expectation $E[b|\hat{v}]$ of the Bias

To obtain an estimate for the conditional bias, the designer can calculate the expectation of the bias conditional on the report, $E[b|\hat{v}] = \int_b b \cdot f(b|\hat{v})\ db$, using the conditional density $f(b|\hat{v})$. For the case of the convolution of two independent uniform random variables, we show in Appendix 7.C that the conditional expectation $E[b|\hat{v}]$ is as follows:

$$E[b|\hat{v}] = \begin{cases} \frac{\hat{v} - \underline{v} + \underline{b}}{2}, & \text{if } \hat{v} < \overline{v} + \underline{b},\ \hat{v} < \underline{v} + \overline{b}, \\ \frac{\hat{v} - \overline{v} + \overline{b}}{2}, & \text{if } \hat{v} \geq \overline{v} + \underline{b},\ \hat{v} \geq \underline{v} + \overline{b}, \\ \frac{2\hat{v} - \underline{v} - \overline{v}}{2}, & \text{if } \hat{v} \geq \overline{v} + \underline{b},\ \hat{v} < \underline{v} + \overline{b}, \\ \frac{\overline{b} + \underline{b}}{2}, & \text{if } \hat{v} < \overline{v} + \underline{b}, \hat{v} \geq \underline{v} + \overline{b}. \end{cases} \tag{5}$$

[8]The entropy of a distribution is a measure of its informativeness. In the case of the uniform distribution $b \sim U\left[\underline{b}, \overline{b}\right]$ it is given by $log(\overline{b} - \underline{b})$ and it is correlated with the variance $\frac{1}{12}(\overline{b} - \underline{b})^2$.

The first row of Equation (5) gives the conditional bias on the lower range from both panels of Figure 1, while the second row gives the conditional bias for the upper range. Row three gives the conditional bias in the middle range when the variance of the true valuation is smaller than the bias ($\overline{v} - \underline{v} < \overline{b} - \underline{b}$) and row four gives the conditional bias in the same range when the opposite holds true ($\overline{v} - \underline{v} \geq \overline{b} - \underline{b}$).

A comparison of the conditional bias with the unconditional bias, $E[b] = (\overline{b} + \underline{b})/2$, reveals that the conditional bias differs from the unconditional bias, which allows the mechanism designer to update her beliefs about the bias of the consumer upon receiving a report $\hat{v}$. The only exception is the middle range, when $\overline{v} - \underline{v} \geq \overline{b} - \underline{b}$. For the lower and upper range, the divergence between the conditional and unconditional bias is intuitive, as very low or very high reports imply low and high biases, respectively. Furthermore, as row three in Equation (5) shows, a mechanism designer can even extract new information on the bias after a report from the middle range, given that the variance of the true valuation is smaller than the variance of the bias ($\overline{v} - \underline{v} < \overline{b} - \underline{b}$). The reason is that these are cases in which $\hat{v}$ can be constituted by any *valuation* $v \in [\underline{v}, \overline{v}]$ and a bias from a "restricted" range $b \in [\hat{v} - \overline{v}, \hat{v} - \underline{v}] \subset [\underline{b}, \overline{b}]$. Only when the opposite holds true, the middle range is uninformative for a mechanism designer and does not allow her to update her beliefs beyond the unconditional mean of the bias. The reason is that these are cases in which $\hat{v}$ can be constituted any *bias* $b \in [\underline{b}, \overline{b}]$ and the according valuation v such that $v + b = \hat{v}$.

5.3 Specific Numerical Examples

We now make further distributional assumptions, discuss optimal non-linear taxation schedules, and contrast them with both, the absence of corrective taxation and the optimal linear tax. Specifically, we consider four scenarios and run a numerical simulation for each of them. Each simulation relies on 100,000 draws. In all scenarios, the true valuation is distributed i.i.d. according to a uniform distribution, $v \sim U[90, 100]$. In the first three scenarios, we furthermore assume that $b \sim U[-10, 10]$, $b \sim U[-5, 5]$, and $b \sim U[-1, 1]$. In all three scenarios, the bias has an expectation of zero, but the variance of the bias is larger than, equal to, and smaller than the variance of the true valuation, respectively. In addition, our fourth scenario considers $b \sim U[-8, 4]$, so that the expectation of the bias is negative (-2) and its variance is exactly as in the first scenario. The results are depicted in the figures of Appendix 8.

A common observation valid for all figures is that the marginal tax rate (Panel (c) in each figure) is lowest at the lowest consumption level, i.e. the lowest perceived type, and highest at the other end of the consumption (and perceived type) distribution. This seems to be juxtaposed to results in many standard mechanism design problems:

in our model the designer should encourage consumption for low types and discourage consumption for high types. The driving force is the functional form of $E[b|\hat{v}]$ (Panel (a) in each figure), which depicts the designer's informational advantage: at the low end of the distribution of $\hat{v}$ she is very certain that consumers are downward biased, and the reverse holds for the upper end of the distribution. This is summarized in the following observation.

Observation 1. *In the uniform case, the marginal tax rate is lowest at the lowest consumption level, and highest at the highest consumption level.*

We proceed by looking at cases, in which the bias is zero in expectation, i.e., Figures 3, 4, and 5 of Appendix 8. Whenever the support of the bias is large compared to that of the true valuation, $\overline{v} - \underline{v} < \overline{b} - \underline{b}$, as for instance in Figure 3, then the optimal non-linear tax yields a qualitatively different allocation than a linear tax (see Panel (b)). This effect becomes smaller as the support of the bias becomes smaller (compare Panels (b) of Figures 3, 4, and 5). As we have mentioned above, in many applications it is plausible to assume that (the support of) the bias is large compared to (that of) the true valuation, so that optimal non-linear taxation would imply quite different allocations than linear taxation in these cases.

The support of the bias in the scenario with uniform distributions critically determines the shape of the functional form of $E[b|\hat{v}]$, which reflects the designer's information advantage: the smaller the variance of the bias, the more the distribution of $E[b|\hat{v}]$ resembles a uniform distribution itself. Since the uniform distribution has maximum entropy, the information advantage decreases in the relative size of the variance of the bias. Obviously, this has an impact on the possibility to correct the Bayesian irrationality: the welfare improvement induced by non-linear taxation decreases as the support of the bias becomes smaller (compare Panels (e), (f), and (g) of Figures 3, 4, and 5). This is summarized in the following observation.

Observation 2. *In the uniform case, the larger the uncertainty about the bias, the larger is the potential for welfare improvement by an optimal non-linear tax scheme compared to no taxation or linear taxation.*

Interestingly, whenever the variance of the bias is larger than that of the true valuation, the optimal mechanism implies bunching contracts for perceived types, which are in the middle range of the distribution. Consider Panel (b) of Figure 3, for instance. Remember that whenever the variance of the bias is larger than that of the true valuation, $\overline{v} - \underline{v} < \overline{b} - \underline{b}$, the designer obtains new information about the expected bias when she receives a report from the middle range of the domain of P. This is not the case if the variance of the bias is smaller than that of the true valuation. Then, reports from the middle range disclose no new information, $E[b|\hat{v}] = E[b]$, see Equation (5). The bunching occurs since the designer's corrective interventions are especially effective on the margins of the domain of P. The middle range, as implicitly defined in Figure 2, is relatively small compared to the margins, whenever the variance of the bias is larger than that of the true valuation. Thus, the designer's optimal mechanism is shaped by her corrective interventions at the margins. This determines the bunching in the middle range.

Observation 3. *In the uniform case, when the uncertainty about the bias is large ($\overline{v} - \underline{v} < \overline{b} - \underline{b}$), the optimal non-linear tax scheme includes bunching contracts for (perceived) types in the middle range.*

In any of the examples depicted in Appendix 8 the highest (perceived) type obtains his optimal allocation in the optimal tax scheme. As in most standard mechnism design problems "there is no distortion at the top". However, additionally, in the behavioral mechanism design setup at hand, the designer is certain that the lowest perceived type would be misoptimizing without her intervention. Thus, due to her informational advantage the designer can construct a mechnism such that also the lowest type obtains his optimal allocation, i.e., "there is no distortion at the bottom". These observations concerning the boundary points of the interval of the perceived valuations determine the overall shape of the optimal mechanism.

Observation 4. *In the uniform case, the optimal non-linear tax scheme implies "no distortion at the top and at the bottom".*

In all scenarios with $E[b] = 0$, linear taxation does not increase welfare compared to no taxation (Panels (e) - (g) of Figures 3, 4, and 5). This is not the case when $E[b] \neq 0$, as in Figure 6: linear taxation increases welfare, since it corrects for the average misoptimization (see Panel (g) of the figure). However, non-linear taxation still improves upon this, since - in addition to correcting for the average misoptimization - it also corrects for the "extreme" misoptimization at the margins of the distribution of the perceived valuation by targeting these consumers with a high degree of Bayesian irrationality. We have already mentioned above that $E[b] < 0$ is a plausible assumption in many real-world applications. Thus, in combination with the observation from above, we see that the designer can notably increase welfare by non-linear taxation in applications with a large bias, that has a negative unconditional expectation.

Observation 5. *In the uniform case, the optimal non-linear tax scheme induces a higher welfare than no taxation and linear taxation.*

Comparing Figures 3 and 6, we can also observe that although the marginal tax rates are qualitatively similar (Panel (c)), the divergence of $E[b]$ from zero implies that the tax rate (Panel (d)) becomes less and less symmetric and for a bias with a purely negative support the tax rate would be monotonically decreasing.

6 Discussion and Conclusion

In this paper, we have derived the optimal non-linear tax to correct internalities, i.e., individual welfare losses from behavioral failures. Using a mechanism design approach, we show that consumers' reports contain information that can be employed to improve upon a linear tax. This beneficial effect of corrective taxation increases in the informativeness of the reports available to the designer.

Empirically, there is evidence that the correlation between reports and biases has some regularities that can be exploited by policy. For example, Allcott et al. (2015) show for energy efficiency investments and hybrid car purchases that higher perceived valuations, i.e., higher reports, are positively correlated with the bias. More specifically, for low perceived valuations, the bias is negative and increases as perceived valuations rise. In such a setting, non-linear taxation should give the largest marginal subsidies to participants with a low perceived valuation. This is also what our analysis suggests.

A positive correlation between reports and biases is also suggested by theories of endogenous inattention (Gabaix, 2014; Kőszegi and Szeidl, 2013; Sallee, 2014). The underlying reason is that participants with large perceived valuations have a larger gross utility gain from learning about biases and will thus pay more attention, leading to a bias closer to zero. Following those arguments, non-linear tax schemes should foresee largest marginal subsidies for participants with low perceived valuations.

To the extent that the positive correlation between reports and bias holds true, many subsidy schemes employed in practice are effectively antipodal to the optimal non-linear tax derived in this paper. For example, the German government grants subsidies for energy efficiency in housing only if a newly built (or retrofitted) house meets predefined minimum efficiencies, so-called "KfW-Effizienzhaus" standards. In other words, marginal subsidies are essentially zero if reports are small and increase only as perceived valuations become larger. As a result, the most heavily biased consumers with low perceived valuations receive no subsidies. This observation implies that social welfare can be improved by implementing the optimal non-linear tax, which specifically targets those consumers.

Since internalities and externalities are often intertwined[9] the inclusion of externalities into our model can be part of future research. Modifying the setup to account

[9]For instance the consumption of highly saccherated soft drinks may cause bad health (internality) and higher burdens for health insurances due to diabetes etc. (externality).

for income effects would be another sensible extension. Moreover, future research could further investigate empirical applications and tests of the model presented in this paper.

7 Appendix: Proofs

7.A Proof of Proposition 1: Optimal Non-Linear Tax

The consumer's first-order condition characterizing optimal consumption x^* is given by

$$\left.\frac{\partial u^d(x, z, \hat{v})}{\partial x}\right|_{x=x^*} = \hat{v} - c'(x^*) - t'(x^*) \overset{!}{=} 0 \Leftrightarrow c'(x^*) = \hat{v} - t'(x^*). \tag{6}$$

The second order condition is satisfied if $-c''(x) - t''(x) \leq 0$ for all $x \in X$. Since the costs are convex by assumption, this condition is satisfied, if the optimal tax schedule is convex as well.[10] Generally, an interior solution exists, if "c is convex enough compared to t", i.e. $c''(x) \geq -t''(x)$ for all $x \in X$.

As discussed in the text we can always guarantee that $\hat{u}\left(\underline{\hat{v}}\right) = \underline{u} \in \mathbb{R}_{++}$ and $\hat{u}\left(\overline{\hat{v}}\right) \geq \underline{u}$, so that the transversality condition immediately implies $\mu\left(\underline{\hat{v}}\right) = 0$ and $\mu\left(\overline{\hat{v}}\right) = 0$. We now use Equation ($\text{FOC}_\text{u}$). By integrating and using $\mu\left(\overline{\hat{v}}\right) = 0$ we obtain

$$\int_{\hat{v}}^{\overline{\hat{v}}} -\mu'(n)dn = -\mu\left(\overline{\hat{v}}\right) - \left[-\mu\left(\hat{v}\right)\right] = \mu\left(\hat{v}\right) \overset{!}{=} \int_{\hat{v}}^{\overline{\hat{v}}} \left(E[\alpha(v)|m] - 1\right) p(m)dm. \tag{7}$$

Using the above equations we rearrange Equation (FOC_x), to obtain the result:

$$\begin{aligned}
&\left(\hat{v} - c'(\cdot)\right) \overset{!}{=} -\frac{\mu\left(\hat{v}\right)}{p(\hat{v})} + E[b|\hat{v}] \cdot E[\alpha(v)|\hat{v}] \\
\overset{(6)}{\Leftrightarrow} \quad & t'(x) = -\frac{\mu\left(\hat{v}_x\right)}{p(\hat{v}_x)} + E[b|\hat{v}_x] \cdot E[\alpha(v)|\hat{v}_x] \\
\overset{(7)}{\Leftrightarrow} \quad & t'(x) = \frac{\int_{\hat{v}_x}^{\overline{\hat{v}}} \left(1 - E[\alpha(v)|m]\right) p(m)dm}{p(\hat{v}_x)} + E[b|\hat{v}_x] \cdot E[\alpha(v)|\hat{v}_x].
\end{aligned}$$

7.B Derivation of the Optimal Linear Tax

In this proof we additionally assume that $c'''(x) = 0$, which simplifies the calculation of the optimal linear tax, but does not change our results on the optimal non-linear tax scheme. Anticipating the behavior on the consumer side, the problem of the designer

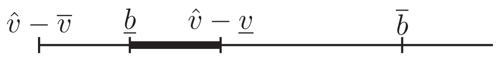
[10] Appendix 8 illustrates that this is the case for most of our examples.

can be written as $\max\limits_{t\in\mathbb{R}} \int_i \int_v \int_b u_i^e(x_i^*)dG(b)dF(v)di + \int_i \int_v \int_b t \cdot x_i^* dG(b)dF(v)di =: V(t)$. We evaluate the derivative with respect to the linear tax t:

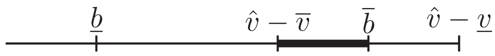
$$\begin{aligned}\frac{\partial V(t)}{\partial t} &= \int_i \int_v \int_b \left[-x_i^* + (v_i - t - c'(\cdot))\frac{\partial x_i^*}{\partial t} \right] dG(b)dF(v)di + \int_i \int_v \int_b \left[x_i^* + t \cdot \frac{\partial x_i^*}{\partial t} \right] dG(b)dF(v)di \\ &= \int_i \int_v \int_b \left[(v_i - c'(\cdot))\frac{\partial x_i^*}{\partial t} \right] dG(b)dF(v)di.\end{aligned}$$

The individually optimal consumption is again characterized by Equation (6), i.e. $c'(\cdot) = \hat{v}_i - t'(x) = (v_i + b_i) - t$, where the last equality holds since $t(x)$ is linear. Thus, $\frac{\partial V}{\partial t} = \int_i \int_v \int_b \left[(t - b_i)\frac{\partial x_i^*}{\partial t} \right] dG(b)dF(v)di$. We can further evaluate $\frac{\partial x_i^*}{\partial t}$ by differentiating Equation (6) with respect to t to obtain $\frac{\partial x_i^*}{\partial t} = -\frac{1}{c''(x_i^*)} = a$ with some constant $a \in \mathbb{R}_{--}$. The last equality follows from the assumption that $c'''(\cdot) = 0$ and that c is convex. Therefore, the optimal tax t^* is given by $\frac{\partial V}{\partial t}|_{t=t^*} \overset{!}{=} 0 \Leftrightarrow t^* = E[b]$.

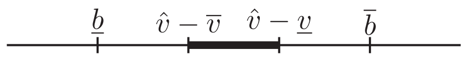
7.C Uniform Case: Illustration of the Calculation of the Conditional Density and of the Conditional Expectation of the Bias



(a) low range:
$\hat{v} < \overline{v} + \underline{b},\ \hat{v} < \underline{v} + \overline{b} \Leftrightarrow \hat{v} - \overline{v} < \underline{b},\ \hat{v} - \underline{v} < \overline{b}$
$\Rightarrow f(b|\hat{v}) = \frac{1}{\hat{v} - \underline{v} - \underline{b}},\ E[b|\hat{v}] = \frac{\hat{v} - \underline{v} + \underline{b}}{2}$



(b) high range:
$\hat{v} \geq \overline{v} + \underline{b},\ \hat{v} \geq \underline{v} + \overline{b} \Leftrightarrow \hat{v} - \overline{v} \geq \underline{b},\ \hat{v} - \underline{v} \geq \overline{b}$
$\Rightarrow f(b|\hat{v}) = \frac{1}{\overline{b} + \overline{v} - \hat{v}},\ E[b|\hat{v}] = \frac{\hat{v} - \overline{v} + \overline{b}}{2}$



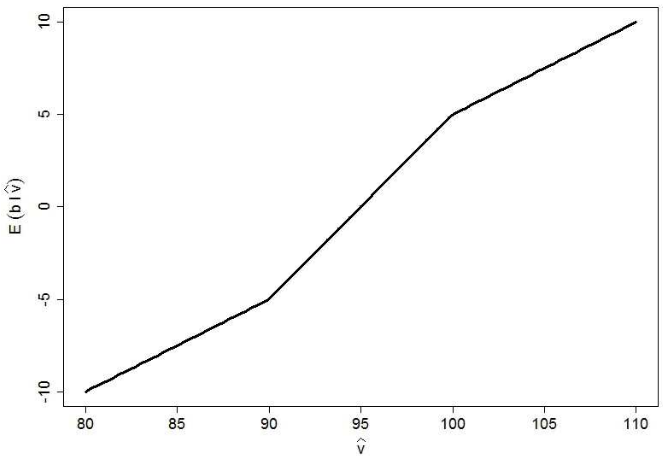
(c) medium range for $\overline{v} - \underline{v} \leq \overline{b} - \underline{b}$:
$\hat{v} \geq \overline{v} + \underline{b},\ \hat{v} < \underline{v} + \overline{b} \Leftrightarrow \hat{v} - \overline{v} \geq \underline{b},\ \hat{v} - \underline{v} < \overline{b}$
$\Rightarrow f(b|\hat{v}) = \frac{1}{\overline{v} - \underline{v}},\ E[b|\hat{v}] = \frac{2\hat{v} - \underline{v} - \overline{v}}{2}$

(d) medium range for $\overline{v} - \underline{v} > \overline{b} - \underline{b}$:
$\hat{v} < \overline{v} + \underline{b},\ \hat{v} \geq \underline{v} + \overline{b} \Leftrightarrow \hat{v} - \overline{v} < \underline{b},\ \hat{v} - \underline{v} \geq \overline{b}$
$\Rightarrow f(b|\hat{v}) = \frac{1}{\overline{b} - \underline{b}},\ E[b|\hat{v}] = \frac{\overline{b} + \underline{b}}{2} = E[b]$

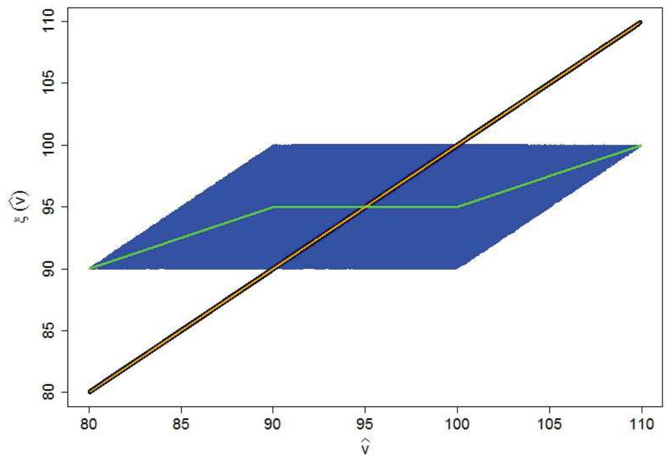
Figure 2: Uniform Case: Conditional density $f(b|\hat{v})$ and conditional expectation $E[b|\hat{v}]$.

8 Appendix: Figures for the Numerical Simulations in Section 5

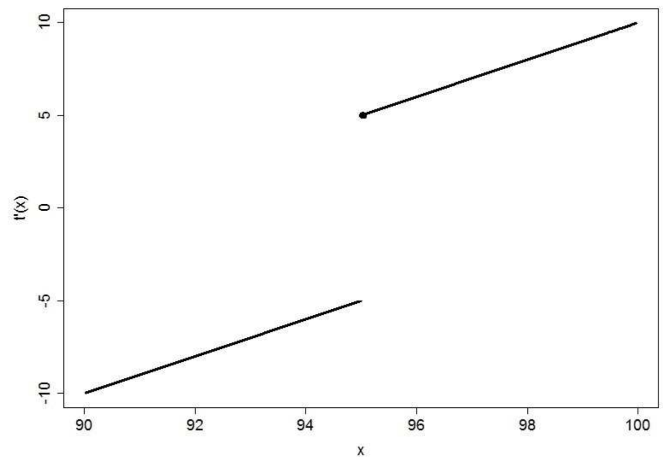
Please turn over! The figures are discussed in Section 5.



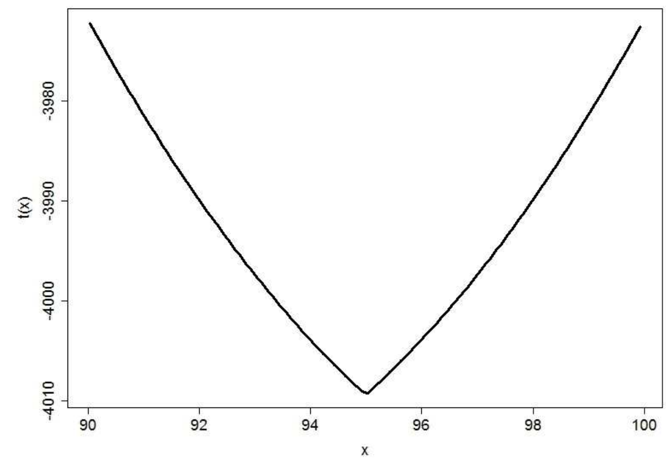
(a) Conditional expectation of the bias $E[b|\hat{v}]$



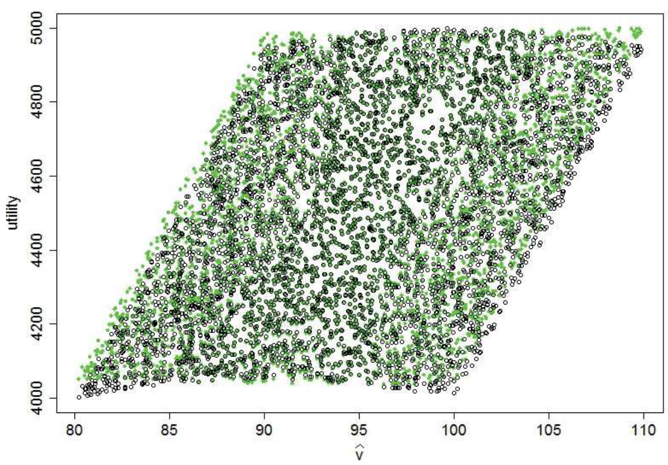
(b) Consumption allocation (black: no tax, blue: optimal [for each possible realization of v, b s.t. $\hat{v} = v + b$], orange: linear tax, green: non-linear tax)



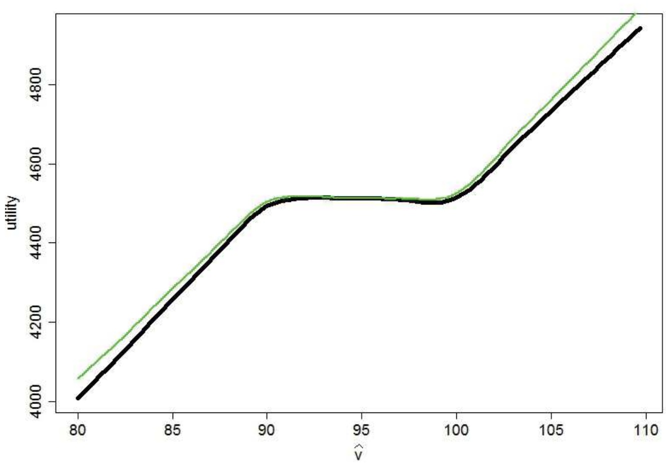
(c) Optimal non-linear marginal tax $t'(x)$



(d) Optimal non-linear tax $t(x)$



(e) Individual experienced utility in equilibrium for each possible realization of v, b s.t. $\hat{v} = v + b$ (black: no tax, green: non-linear tax)

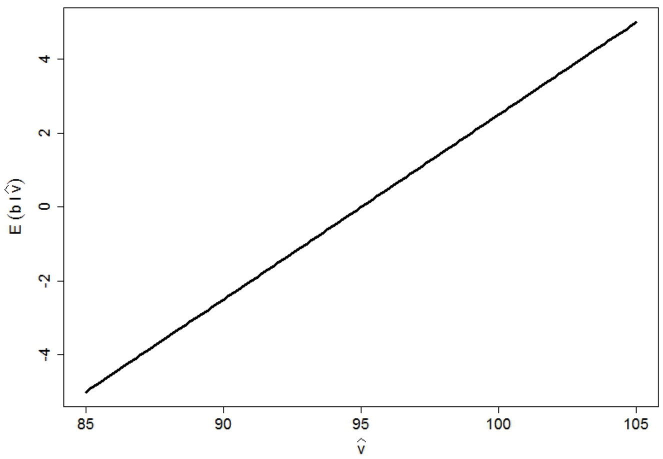


(f) Expected experienced utility in equilibrium (black: no tax, green: non-linear tax)

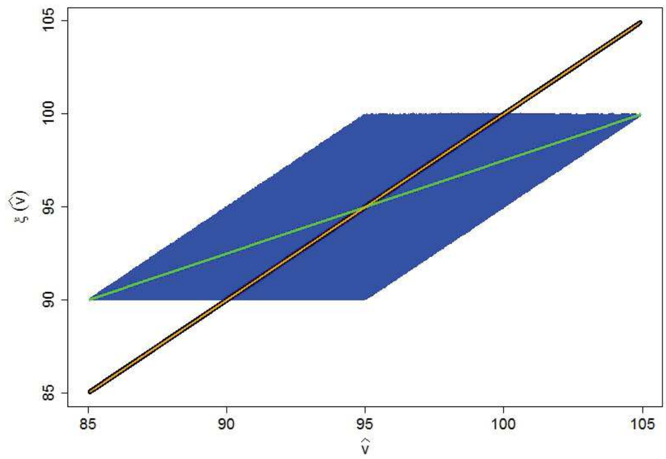
no tax: 4500.528 | linear tax: 4500.528 | non-linear tax: 4514.045

(g) Expected total welfare

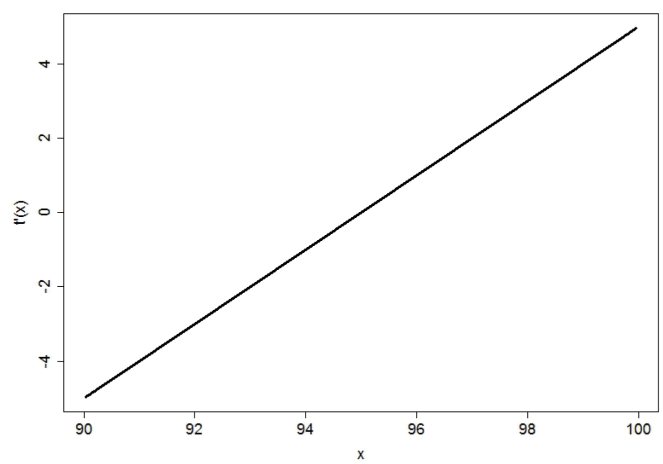
Figure 3: Scenario with $v \sim U[90, 100]$ and $b \sim U[-10, 10]$, i.e. $E[b] = 0$.



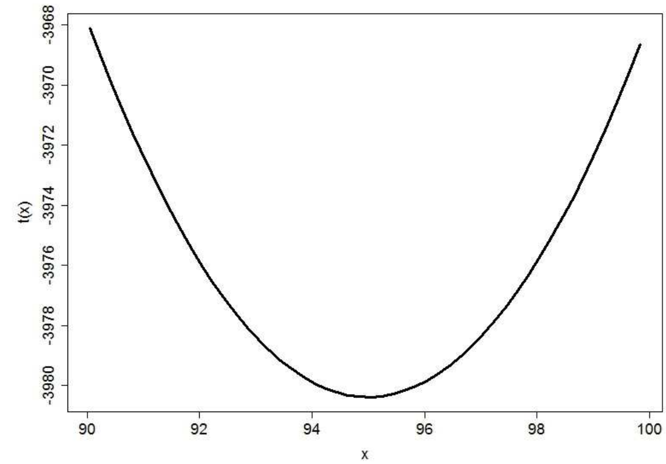
(a) Conditional expectation of the bias $E[b|\hat{v}]$



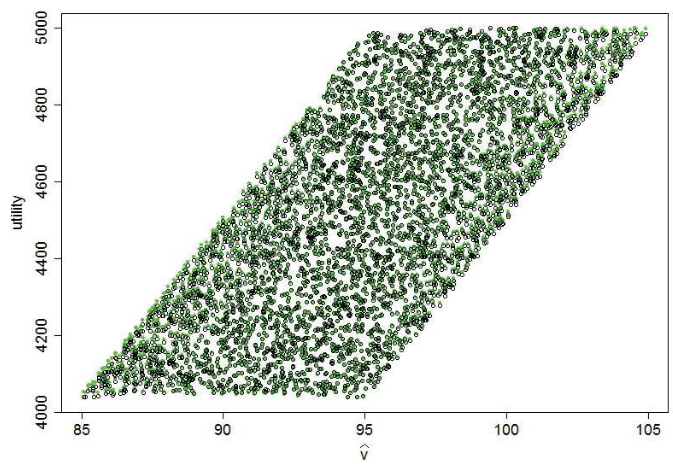
(b) Consumption allocation (black: no tax, blue: optimal [for each possible realization of v, b s.t. $\hat{v} = v + b$], orange: linear tax, green: non-linear tax)



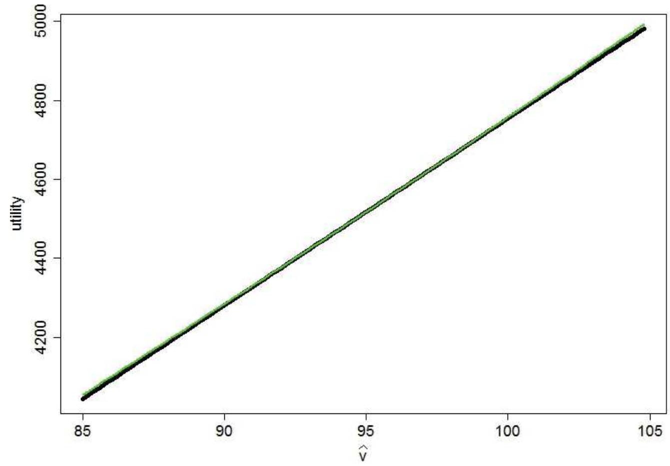
(c) Optimal non-linear marginal tax $t'(x)$



(d) Optimal non-linear tax $t(x)$



(e) Individual experienced utility in equilibrium for each possible realization of v, b s.t. $\hat{v} = v + b$ (black: no tax, green: non-linear tax)

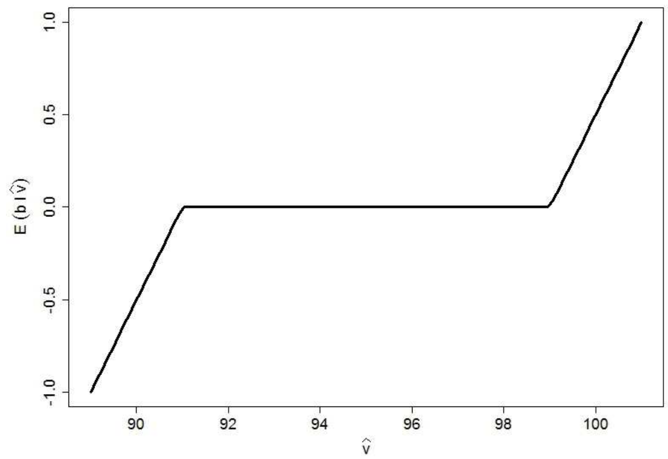


(f) Expected experienced utility in equilibrium (black: no tax, green: non-linear tax)

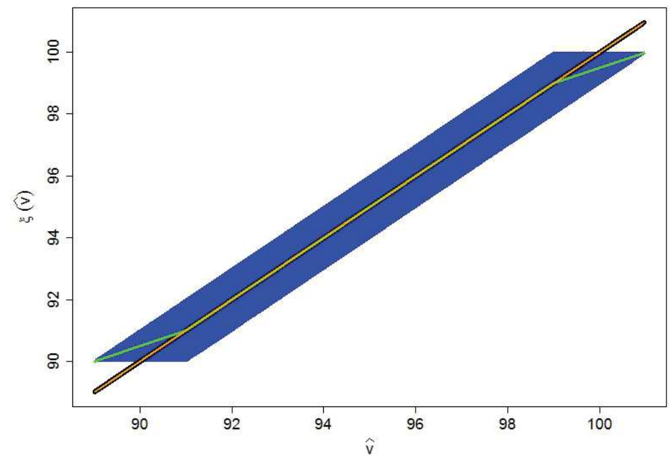
no tax: 4513.158 | linear tax: 4513.158 | non-linear tax: 4515.226

(g) Expected total welfare

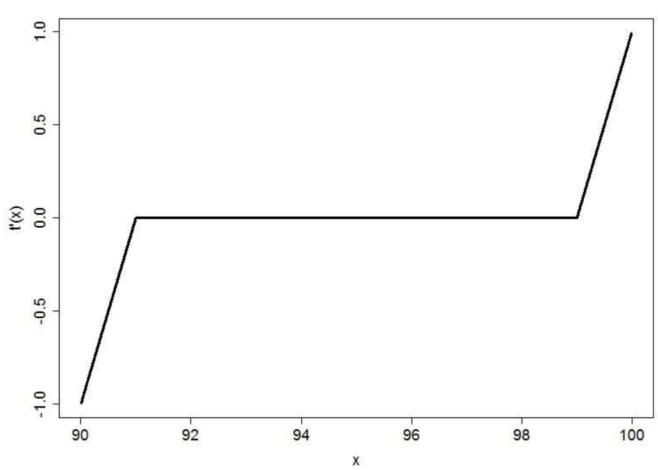
Figure 4: Scenario with $v \sim U[90, 100]$ and $b \sim U[-5, 5]$, i.e. $E[b] = 0$.



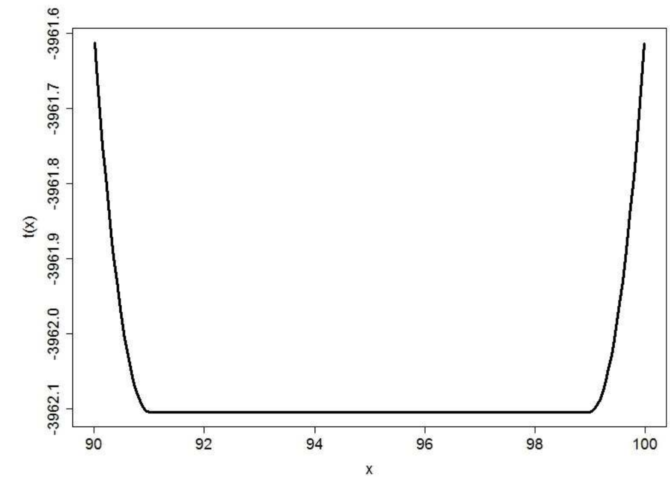
(a) Conditional expectation of the bias $E[b|\hat{v}]$



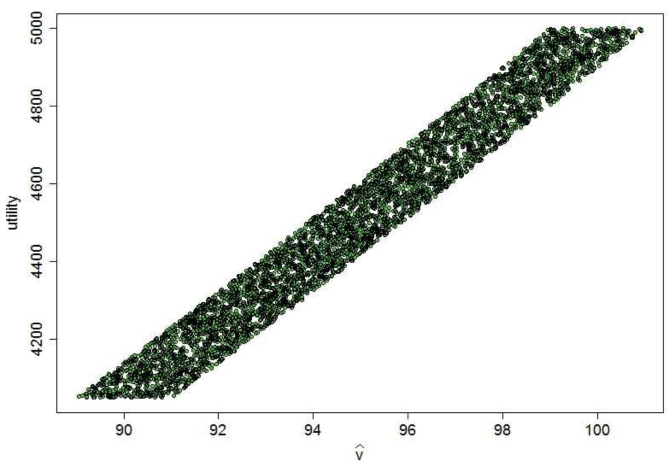
(b) Consumption allocation (black: no tax, blue: optimal [for each possible realization of v, b s.t. $\hat{v} = v + b$], orange: linear tax, green: non-linear tax)



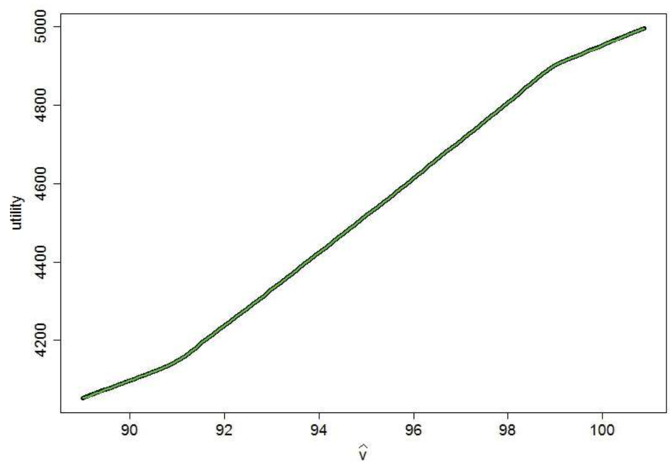
(c) Optimal non-linear marginal tax $t'(x)$



(d) Optimal non-linear tax $t(x)$



(e) Individual experienced utility in equilibrium for each possible realization of v, b s.t. $\hat{v} = v + b$ (black: no tax, green: non-linear tax)

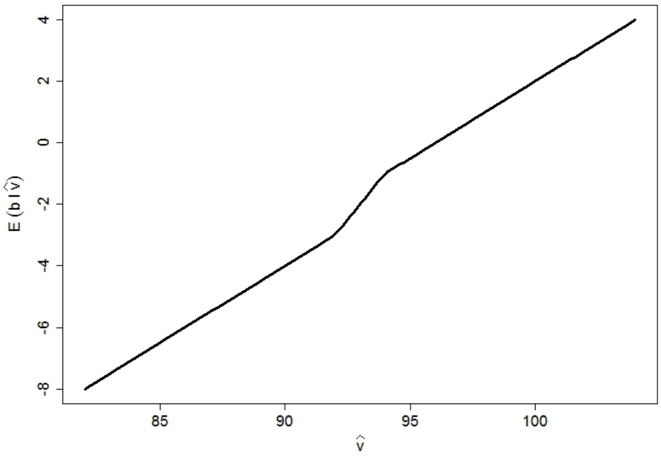


(f) Expected experienced utility in equilibrium (black: no tax, green: non-linear tax)

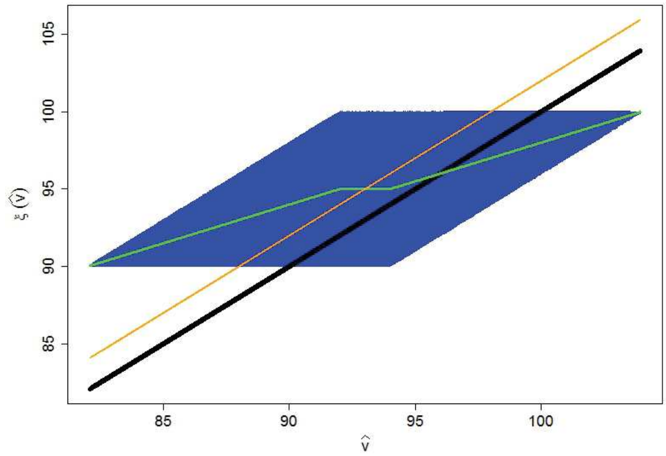
no tax: 4515.744 | linear tax: 4515.744 | non-linear tax: 4515.761

(g) Expected total welfare

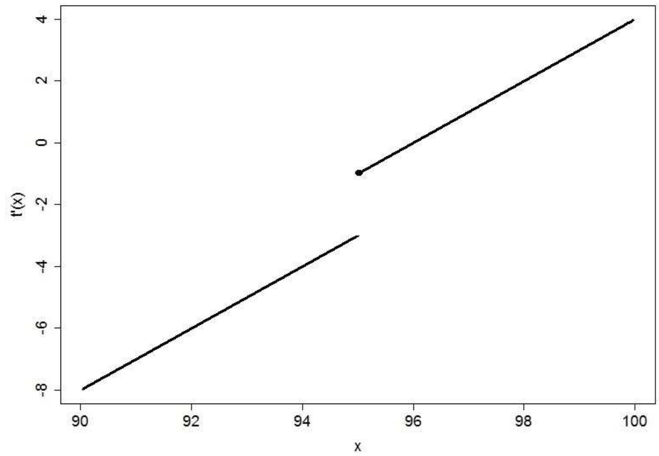
Figure 5: Scenario with $v \sim U[90, 100]$ and $b \sim U[-1, 1]$, i.e. $E[b] = 0$.



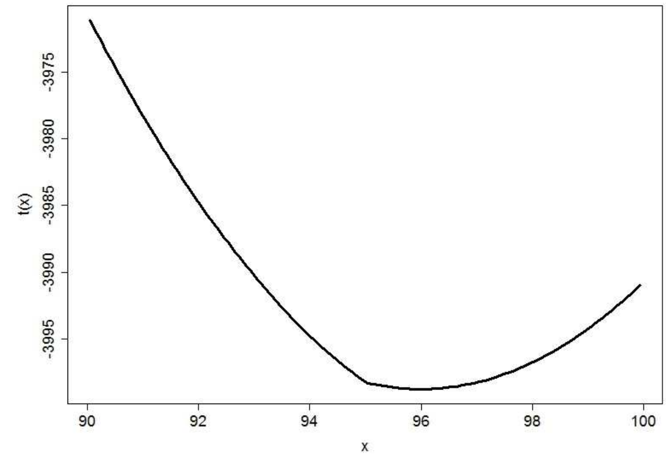
(a) Conditional expectation of the bias $E[b|\hat{v}]$



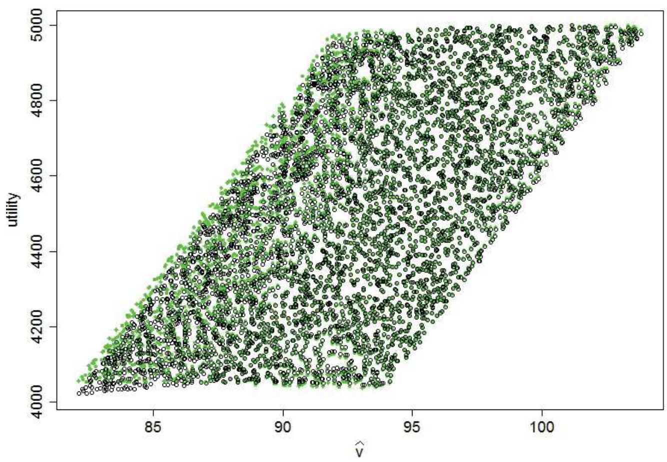
(b) Consumption allocation (black: no tax, blue: optimal [for each possible realization of v, b s.t. $\hat{v} = v + b$], orange: linear tax, green: non-linear tax)



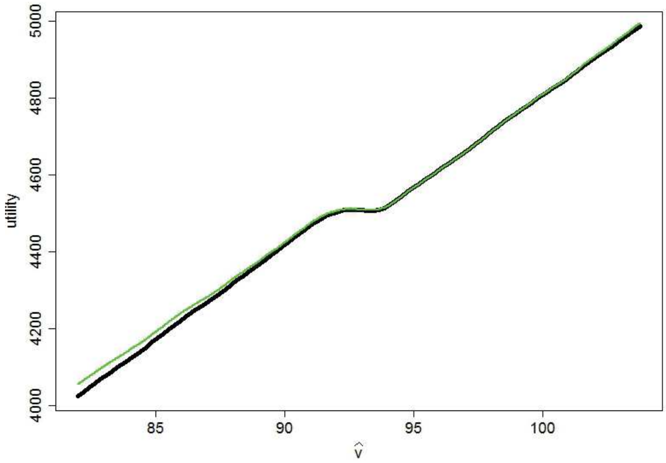
(c) Optimal non-linear marginal tax $t'(x)$



(d) Optimal non-linear tax $t(x)$



(e) Individual experienced utility in equilibrium for each possible realization of v, b s.t. $\hat{v} = v + b$ (black: no tax, green: non-linear tax)



(f) Expected experienced utility in equilibrium (black: no tax, green: non-linear tax)

no tax: 4507.835 | linear tax: 4509.821 | non-linear tax: 4513.393

(g) Expected total welfare

Figure 6: Scenario with $v \sim U[90, 100]$ and $b \sim U[-8, 4]$, i.e. $E[b] = -2$.

References

Abaluck, Jason and Jonathan Gruber, Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program, *The American Economic Review*, 2011, *101*, 1180–1210.

Akerlof, George A., The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning, *The American Economic Review*, 1978, *68*, 8–19.

Allcott, Hunt, Paternalism and Energy Efficiency: An Overview, *Annual Review of Economics*, 2016, *8*, 145–176.

Allcott, Hunt and Dmitry Taubinsky, Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market, *The American Economic Review*, 2015, *105*, 2501–2538.

Allcott, Hunt, Christopher Knittel, and Dmitry Taubinsky, Tagging and Targeting of Energy Efficiency Subsidies, *The American Economic Review*, 2015, *105*, 187–191.

Allcott, Hunt, Sendhil Mullainathan, and Dimitry Taubinsky, Energy Policy with Externalites and Internalities, *Journal of Public Economics*, 2014, pp. 72–88.

Allcott, Hunt, Sendhil Mullainathan, Meredith Fowlie, Kenneth Gillingham, Danny Goroff, Matt Harding, Ted Howes, Kelsey Jack, Paul L. Joskow, Ogi Kavazovic, Tony Leiserowitz, Dave Rapson, Todd Rogers, Eldar Shafir, Rob Stavins, and Gernot Wagner, Behavioral Science and Energy Policy, 2017. Working Paper.

Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and Wändi Bruine de Bruin, Public Perceptions of Energy Consumption and Savings, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, *107*, 16054–16059.

Barlow, Richard E., Albert W. Marshall, and Frank Proschan, Properties of Probability Distributions with Monotone Hazard Rate, *The Annals of Mathematical Statistics*, 1963, *34*, 375–389.

Bertsekas, Dimitri P. and John N. Tsitsiklis, *Introduction to Probability*, Athena Scientific, 2008.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, Calorie Posting in Chain Restaurants, *American Economic Journal: Economic Policy*, 2011, *3*, 91–128.

Börgers, Tilman, *An Introduction in the Theory of Mechanism Design*, Oxford University Press, 2015.

Camerer, Colin F., *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, 2003.

Chetty, Raj, Behavioral Economics and Public Policy: A Pragmatic Perspective, *The American Economic Review: Papers and Proceedings*, 2015, *105*, 1–33.

Congdon, William J., Jeffrey R. Kling, and Sendhil Mullainathan, *Policy and Choice: Public Finance through the Lens of Behavioral Economics*, Brookings Institution Press, 2011.

Cremer, Helmuth and Firouz Gahvari, On Optimal Taxation of Housing, *Journal of Urban Economics*, 1998, *43*, 315–335.

Cremer, Helmuth, Firouz Gahvari, and Norbert Ladoux, Externalities and Optimal Taxation, *Journal of Public Economics*, 1998, *70*, 343–364.

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

Farhi, Emmanuel and Xavier Gabaix, Optimal Taxation with Behavioral Agents, 2017. Working Paper.

Gabaix, Xavier, A Sparsity-Based Model of Bounded Rationality, *The Quarterly Journal of Economics*, 2014, *129*, 1661–1710.

Gerritsen, Aart, Optimal Taxation When People Do Not Maximize Well-being, *Journal of Public Economics*, 2016, *144*, 122–139.

Jensen, Robert, The (Perceived) Returns to Education and the Demand for Schooling, *The Quarterly Journal of Economics*, 2010, *125*, 515–548.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin, Back to Bentham? Explorations of Experienced Utility, *The Quarterly Journal of Economics*, 1997, *112*, 375–405.

Kőszegi, Botond and Adam Szeidl, A Model of Focusing in Economic Choice, *The Quarterly Journal of Economics*, 2013, *128*, 53–104.

Kőszegi, Botond, Behavioral Contract Theory, 2014. Working Paper.

Laibson, David, Golden Eggs and Hyperbolic Discounting, *The Quarterly Journal of Economics*, 1997, *112*, 443–478.

Lindahl, Erik, *Die Gerechtigkeit der Besteuerung*, Hakan Ohlssons Buchdruckerei, 1919.

Lockwood, Benjamin B. and Dmitry Taubinsky, Optimal Income Taxation with Present Bias, 2016. Working Paper.

Lockwood, Benjamin B. and Dmitry Taubinsky, Regressive Sin Taxes, 2017. Working Paper.

Loewenstein, George and Drazen Prelec, Anomalies in Intertemporal Choice: Evidence and an Interpretation, *The Quarterly Journal of Economics*, 1992, *107* (2), 573–597.

Mirrlees, James A., An Exploration in the Theory of Optimum Income Taxation, *The Review of Economic Studies*, 1971, *38*, 175–208.

Mirrlees, James A., Optimal Tax Theory: A Synthesis, *Journal of Public Economics*, 1976, *6*, 327–358.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon, A Reduced-Form Approach to Behavioral Public Finance, *Annual Review of Economics*, 2012, *4*, 511–540.

Mussa, Michael and Sherwin Rosen, Monopoly and Product Quality, *Journal of Economic Theory*, 1978, *18*, 301–317.

O'Donoghue, Ted and Matthew Rabin, Doing It Now or Later, *The American Economic Review*, 1999, *89*, 103–124.

O'Donoghue, Ted and Matthew Rabin, Optimal Sin Taxes, *Journal of Public Economics*, 2006, *90*, 1825–1849.

Oliver, Adam, *Behavioural Public Policy*, Cambridge University Press, 2013.

Pigou, Arthur C., *The Economics of Welfare*, Macmillan, 1920.

Ramsey, Frank, A Contribution to the Theory of Taxation, *Economic Journal*, 1927, *37*, 47–61.

Rees-Jones, Alex and Dmitry Taubinsky, Taxing Humans: Pitfalls of the Mechanism Design Approach and Potential Resolutions, 2017. Working Paper.

Saez, Emmanuel, Using Elasticities to Derive Optimal Income Tax Rates, *The Review of Economic Studies*, 2001, *68*, 205–229.

Salanié, Bernard, *The Economics of Taxation*, The MIT Press, 2011.

Sallee, James M., Rational Inattention and Energy Efficiency, *The Journal of Law and Economics*, 2014, *57*, 781–820.

Sunstein, Cass C., *The Ethics of Influence: Government in the Age of Behavioral Science*, Cambridge University Press, 2016.

Sydsaeter, Knut, Peter Hammond, Alte Seierstad, and Arne Strom, *Further Mathematics for Economic Analysis*, Pearson Education, 2008.

Wilson, Robert B., *Nonlinear Pricing*, Oxford University Press, 1997.