Online Appendix

Optimal Internality Taxation of Product Attributes

Andreas Gerster and Michael Kramm

A Proofs

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

We first use the mechanism design approach to characterize the optimal tax in terms of types and then employ the perturbation approach to characterize the optimal tax as a function of the attribute level. Finally, we show that the results we obtain under either approach are equivalent.

A.A.1 Equation (2) in Types (Mechanism Design Approach)

For our mechanism design approach, we employ the Revelation Principle for dominant-strategy implementation (Gibbard, 1973) to solve for a direct mechanism, where the consumer truthfully reveals information about his perceived valuation.

We start by discussing the dimensionality of mechanisms for internality taxation. In our model, a consumer decides based on his perceived valuation $\hat{v}(v, b)$ rather than v. Even though a consumer may or may not know his bias, i.e., be sophisticated or naive, we can, without loss of generality, neglect that distinction and restrict our analysis to one-dimensional mechanisms in the perceived valuation $\hat{v}$. Naive consumers are unaware of their bias and thus cannot report it, so that a social planner can only employ a one-dimensional mechanism in $\hat{v}$ to correct them. Sophisticated consumers know their bias and can, in principle, report it. Yet, as biases do not influence decision utility, truth-telling in a two-dimensional mechanism is not incentive-compatible.

The formal argument of the above is as follows. We want to show that two-dimensional mechanisms, where sophisticated consumers report both their perceived valuation and their bias, lead to a violation of truth-telling. Without loss of generality, we assume that there exists at least one realization of perceived valuations $\hat{v}_1 \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ for which biases differ, so that some consumers are characterized by $(\hat{v}_1, b_1)$ and others by $(\hat{v}_1, b_0)$, where $b_1 \neq b_0$.

The policy maker wants to implement a direct two-dimensional mechanism where the allocation $\xi\left(\tilde{v},\tilde{b}\right)$ and the tax $\tau\left(\tilde{v},\tilde{b}\right)$ depend on reported valuations $\tilde{v}$ and biases $\tilde{b}$. Consumers choose their reports to maximize decision utility:

$$(\tilde{v}^*(\hat{v}),\tilde{b}^*(\hat{v})) = \arg\max_{\tilde{v},\tilde{b}} u^d\left(\xi\left(\tilde{v},\tilde{b}\right),\tau\left(\tilde{v},\tilde{b}\right)|\hat{v}\right).$$

Importantly, this maximization problem depends only on reported biases $\tilde{b}$ and not on actual biases b. As a consequence, every sophisticated consumer with perceived valuation $\hat{v}$ will report the same bias $\tilde{b}^*(\hat{v})$. As biases differ for $\hat{v}_1$, truth-telling is violated.

Hence, we can apply the Revelation Principle to our setting, where the space of reports for a consumer is given by the space of his perceived valuation $\hat{v}$. For simplicity, we refer to $\hat{v}$ as a consumer's type in the following. The policy maker confines herself to designing a direct mechanism $(\xi,\tau) : \left[\underline{\hat{v}},\overline{\hat{v}}\right] \to Q \times \mathbb{R}$ under truth-telling in order to implement the welfare maximizing outcome. Based on the consumer's strategical report $\tilde{v}$, the allocation rule of the direct mechanism assigns the consumed attribute level, $\xi(\tilde{v}) \in Q$, and transfer rule the amount of taxes to be paid, $\tau(\tilde{v}) \in \mathbb{R}$. Because the consumer has unit demand for the good, participation constraints are not relevant in our setting.

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

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

Since the consumer may strategically misreport his perceived valuation, truth-telling must be ensured by implementing an incentive compatible mechanism. This implies that the tax schedule must satisfy:

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

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

$$\hat{v}\xi'(v^*) - \tau'(v^*) - \xi'(v^*)c'(\xi(v^*)) \stackrel{!}{=} 0. \tag{8}$$

As incentive compatibility requires that $v^* = \hat{v}$, equilibrium decision utility in an incentive-compatible direct mechanism is given by $\hat{u}^d(\hat{v}) := u^d(\xi(\hat{v}),\tau(\hat{v})|\hat{v})$, while equilibrium normative utility is given by $\hat{u}^n(\hat{v},b) := u^n(\xi(\hat{v}),\tau(\hat{v})|v) = \hat{u}^d(\hat{v}) - b\xi(\hat{v})$.

Put differently, incentive compatibility implies that, for all $\hat{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$, the following equation has to hold:

$$\frac{\partial \hat{u}^d(\hat{v})}{\partial \hat{v}} = \xi(\hat{v}) + \hat{v}\xi'(\hat{v}) - \tau'(\hat{v}) - \xi'(\hat{v})c'(\xi(\hat{v})) \overset{(8)}{=} \xi(\hat{v}). \tag{9}$$

To determine the optimal tax schedule, the policy maker solves an optimization problem which can be analyzed using an optimal control approach. 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 policy maker is given by:

$$\max_{\xi \in \mathbb{Q}} \int_{\hat{v}} \alpha(\hat{v}) \cdot E\left[\hat{u}^n\left(\hat{v}, b\right) | \hat{v}\right] dF(\hat{v}) + \lambda \left(\int_{\hat{v}} \tau\left(\hat{v}\right) dF(\hat{v}) - B \right), \tag{10}$$

subject to the condition from Equation (9), where $\mathbb{Q} := \{f | f : \left[\underline{\hat{v}}, \overline{\hat{v}}\right] \to Q\}$ is the function space containing all functions with domain $\left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ and codomain Q. 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 determined by incentive compatibility and optimal strategic reporting, as given by Equation (9).

Using the definition of decision utility to replace the tax and rewriting equilibrium normative utility in terms of equilibrium decision utility, the Hamiltonian for the problem stated in Equation (10) for all $\hat{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]$ is given by:

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

Following the standard solution procedure for such mechanism design problems, we employ Pontryagin's Maximum Principle, which yields the following necessary conditions for the optimal tax.[20]

$$\text{FOC on control: } \frac{\partial \mathfrak{H}}{\partial \xi} = \left[-E[b|\hat{v}] \cdot \alpha(\hat{v}) + \lambda\left(\hat{v} - c'(\cdot)\right)\right] f(\hat{v}) + \mu\left(\hat{v}\right) \overset{!}{=} 0, \tag{FOC$_q$}$$

$$\text{FOC on state: } \frac{\partial \mathfrak{H}}{\partial \hat{u}^d} = \left[\alpha(\hat{v}) - \lambda\right] f(\hat{v}) \overset{!}{=} -\mu'\left(\hat{v}\right), \tag{FOC$_u$}$$

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

[20] In addition, sufficiency is given if the control region is convex and the Hamiltonian is concave in $(\xi, \hat{u}^d)$ for every $\hat{v}$. Both conditions are satisfied in our setup.

The consumer's first-order condition characterizing optimal consumption q^d is given by

$$\left.\frac{\partial u^d(q,t,\hat{v})}{\partial q}\right|_{q=q^d} = \hat{v} - c'(q^d) - t'(q^d) \stackrel{!}{=} 0 \Leftrightarrow c'(q^d) = \hat{v} - t'(q^d). \tag{11}$$

The second order condition is satisfied if $c''(q) + t''(q) \geq 0$ for all $q \in Q$. Since the costs are convex in q by assumption, this condition is satisfied if the optimal tax schedule is convex in q as well. We find that convex optimal tax schedules are optimal for many behavioral biases (see discussion after Proposition 4). More generally, this condition also holds if the optimal tax schedule is concave. In that case, the cost function may not be "too concave" in the sense that $c''(q) + t''(q) \geq 0$ holds.

In our setting, we abstract from participation constraints. Hence, we can w.l.o.g. assume that $\hat{u}\left(\underline{\hat{v}}\right) = \underline{u} > 0$ 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 $\mu\left(\overline{\hat{v}}\right) = 0$. Integrating Equation ($\text{FOC}_\text{u}$) and using $\mu\left(\overline{\hat{v}}\right) = 0$, we obtain

$$\int_{\hat{v}}^{\overline{\hat{v}}} -\mu'(n)dn = -\mu\left(\overline{\hat{v}}\right) - \left[-\mu\left(\hat{v}\right)\right] = \mu\left(\hat{v}\right) \stackrel{!}{=} \int_{\hat{v}}^{\overline{\hat{v}}} \left[\alpha(m) - \lambda\right] f(m)dm. \tag{12}$$

Using the Equations (11) and (12), we rearrange Equation (FOC_q), to obtain the result:

$$\begin{aligned}
\lambda\left(\hat{v} - c'(\cdot)\right) &\stackrel{!}{=} -\frac{\mu\left(\hat{v}\right)}{f(\hat{v})} + E[b|\hat{v}] \cdot \alpha(\hat{v}) \\
\stackrel{(11)}{\Leftrightarrow} \quad t'(q) &= -\frac{\mu\left(\hat{v}_q\right)}{\lambda f(\hat{v}_q)} + E[b|\hat{v}_q] \cdot \frac{\alpha(\hat{v}_q)}{\lambda} \\
\stackrel{(12)}{\Leftrightarrow} \quad t'(q) &= \frac{\int_{\hat{v}_q}^{\overline{\hat{v}}} \left[1 - \hat{g}(m)\right] f(m)dm}{f(\hat{v}_q)} + E[b|\hat{v}_q] \cdot \hat{g}(\hat{v}_q).
\end{aligned}$$

The average marginal social welfare weight for consumers in the upper part of the distribution (in comparison to $\hat{v}$) is defined as:

$$\hat{G}(\hat{v}) := \frac{\int_{\hat{v}}^{\overline{\hat{v}}} \hat{g}(m)dF(m)}{1 - F(\hat{v})} = E[g(\theta)|\theta \geq \hat{v}]. \tag{13}$$

Using this we get

$$\begin{aligned}
t'(q_{\hat{v}}) &= \hat{g}(\hat{v})E\left[b|\hat{v}\right] + \frac{\int_{\hat{v}}^{\overline{\hat{v}}} \left[1 - \hat{g}(n)\right] f(n)dn}{f(\hat{v})} \\
&= \hat{g}(\hat{v})E\left[b|\hat{v}\right] + \frac{\int_{\hat{v}}^{\overline{\hat{v}}} f(n)dn - \int_{\hat{v}}^{\overline{\hat{v}}} \hat{g}(n)f(n)dn}{f(\hat{v})}
\end{aligned}$$

$$\overset{(13)}{=} \hat{g}(\hat{v})E\left[b|\hat{v}\right] + \frac{\left[1-F(\hat{v})\right] - \hat{G}(\hat{v})\left[1-F(\hat{v})\right]}{f(\hat{v})}$$
$$= \hat{g}(\hat{v})E\left[b|\hat{v}\right] + \frac{\left[1-F(\hat{v})\right]\left[1-\hat{G}(\hat{v})\right]}{f(\hat{v})}$$

A.A.2 Equation (3) in Attribute Level q (Perturbation Approach)

The perturbation approach uses insights from a local perturbation of the tax schedule to derive the optimal tax. In particular, it examines a change in the slope of the consumer's budget set in a small band $(q, q+dq)$ from $1-c'(q)-t'(q)$ to $1-c'(q)-t'(q)-d\tau$. Such a change has four effects on welfare. The mechanical effect represents an increase in collected tax money, since all consumers with consumption levels higher than $q+dq$ are affected by the increased tax rate. This effect is captured by

$$d\tau dq\left[1-H(q)\right].$$

The consumer welfare effect captures the social value of the decrease in consumption of the numeraire due to the local tax perturbation:

$$-d\tau dq\left[1-H(q)\right]G(q).$$

Since we abstract from income effects, there is no change of the consumption of q for these consumers. This is different for the consumers inside the small band $(q, q+dq)$. Here, relative prices are changed and thus consumers change their consumption of the attribute level by

$$\delta q = \frac{dq}{d\left(c'+t'\right)}d\tau = -e\frac{q}{c'+t'}d\tau. \tag{14}$$

where the elasticity is defined as

$$e := -\frac{dq}{d\left(t'+c'\right)}\frac{t'+c'}{q}. \tag{15}$$

We define the elasticity at the net price $t'+c'$, since the slope of the consumer's budget constraint is determined by $t'+c'$.

The fiscal externality captures the decrease in collected taxes

$$\delta q t'(q)h(q)dq,$$

while the bias correction effect

$$-g(q)\delta q E\left[b|q\right]h(q)dq$$

captures the change in welfare due to the induced change in consumer utility, which is caused by the potential correction of an internality.

In optimum, a local perturbation of the tax may not change social welfare. Hence, the four effects have to sum to zero, so that

$$\begin{aligned}
0 &= d\tau dq\left[1-H(q)\right]\left[1-G(q)\right]+\delta q\left[t'(q)-E\left[b|q\right]g(q)\right]h(q)dq \\
\Leftrightarrow t'(q) &= g(q)E\left[b|q\right]-\frac{d\tau\left[1-H(q)\right]\left[1-G(q)\right]}{\delta q h(q)} \\
&\overset{(14)}{=} g(q)E\left[b|q\right]-\frac{d\tau\left[1-H(q)\right]\left[1-G(q)\right]}{-e(q)\frac{q}{t'(q)+c'(q)}d\tau h(q)} \\
&= g(q)E\left[b|q\right]+\frac{1-G(q)}{e(q)a(q)}\left(t'(q)+c'(q)\right).
\end{aligned}$$

A.A.3 Equivalence of Equations (2) and (3)

We first linearize the budget constraint of the consumer in the tax so that, using the virtual income R, it can be rewritten as $z = R - c(q) - \tau q$. Then, the decision utility is given by $u = \hat{v}q + R - c(q) - \tau q$.

The goal is now to write the change of consumption q in type $\hat{v}$ using elasticity e, which is the elasticity of q with respect to net-price $\tau + c'$. The first-order condition of the consumer with respect to consumption gives

$$0 \overset{!}{=} \hat{v} - \tau - c'(q). \tag{16}$$

Applying the implicit function theorem twice to (16) yields

$$\begin{aligned}
\frac{dq}{d\hat{v}} &= -\frac{1}{-c''(q)} = \frac{1}{c''(q)} \\
\frac{dq}{d\left(\tau+c'\right)} &= -\frac{-1}{-c''(q)} = -\frac{1}{c''(q)} \\
\Rightarrow \frac{dq}{d\hat{v}} &= -\frac{dq}{d\left(\tau+c'\right)}.
\end{aligned} \tag{17}$$

We use the above results and the definition of the elasticity from (15):

$$e = -\frac{dq}{d\left(\tau+c'\right)}\frac{\tau+c'}{q}$$

$$\overset{(17)}{=} \frac{dq}{d\hat{v}} \frac{\tau + c'}{q}$$
$$\Leftrightarrow \frac{dq}{d\hat{v}} = \frac{qe}{\tau + c'}. \tag{18}$$

For density $h(q)$ the following must hold

$$h(q_{\hat{v}})dq_{\hat{v}} = f(\hat{v})d\hat{v}$$
$$\Leftrightarrow h(q_{\hat{v}})\frac{dq(\hat{v})}{d\hat{v}} = f(\hat{v})$$
$$\overset{(18)}{\Leftrightarrow} h(q_{\hat{v}})\frac{qe}{t' + c'} = f(\hat{v}). \tag{19}$$

The average marginal social welfare weight for consumers in the upper part of the distribution (in comparison to q) can be rewritten using a change of variables (Ch.o.V.):

$$\hat{G}(\hat{v}) := \frac{\int_{\hat{v}}^{\overline{\hat{v}}} \hat{g}(m)dF(m)}{1 - F(\hat{v})} \overset{Ch.o.V}{=} \frac{\int_{q_{\hat{v}}}^{\overline{q}} g(m)dH(q_m)}{1 - H(q_{\hat{v}})} =: G(q_{\hat{v}}). \tag{20}$$

The thinness of the top tail (individuals with consumption above q in relation to those at q) is given by

$$a(q) := \frac{h(q)q}{1 - H(q)}. \tag{21}$$

Using the above results we get

$$t'(q_{\hat{v}}) = \hat{g}(\hat{v})E\left[b|\hat{v}\right] + \frac{\left[1 - F(\hat{v})\right]\left[1 - \hat{G}(\hat{v})\right]}{f(\hat{v})}$$
$$\overset{(19),C.o.V.}{=} g(q)E\left[b|q\right] + \frac{\left[1 - H(q)\right]\left[1 - G(q)\right]}{h(q)\frac{qe(q)}{t'(q) + c'(q)}}$$
$$= g(q)E\left[b|q\right] + \frac{\left(t'(q) + c'(q)\right)\left[1 - H(q)\right]\left[1 - G(q)\right]}{h(q)qe(q)}$$
$$\Leftrightarrow t'(q) \overset{(21)}{=} g(q)E\left[b|q\right] + \frac{1 - G(q)}{a(q)e(q)}\left(t'(q) + c'(q)\right).$$

A.B Derivation of the Optimal Linear Tax

The policy maker's problem of setting the optimal linear tax can be written as

$$\max_{t \in \mathbb{R}} \int_v \int_b \alpha(v, b)u^n(q^d, t, v)dF_b(b|v)dF_v(v) + \lambda\left(\int_v \int_b t \cdot q^d dF_b(b|v)dF_v(v) - B\right) =: V(t).$$

We evaluate the derivative with respect to the linear tax t:

$$\frac{\partial V(t)}{\partial t} = \int_v \int_b \alpha(v,b) \left[-q^d + \big(v - t - c'(\cdot)\big) \frac{\partial q^d}{\partial t} \right] dF_b(b|v) dF_v(v) + \lambda \int_v \int_b \left[q^d + t \cdot \frac{\partial q^d}{\partial t} \right] dF_b(b|v) dF_v(v)$$
$$= \int_v \int_b \left[\alpha(v,b) \big(v - c'(\cdot)\big) \frac{\partial q^d}{\partial t} - (\alpha(v,b) - \lambda)(q^d + t \frac{\partial q^d}{\partial t}) \right] dF_b(b|v) dF_v(v).$$

The individually optimal consumption is again characterized by Equation (11), i.e., $c'(\cdot) = \hat{v} - t'(q) = (v + b) - t$, where the last equality holds since t is linear. Thus,

$$\frac{\partial V}{\partial t} = \int_v \int_b \left[\alpha(v,b)(t - b) \frac{\partial q^d}{\partial t} - (\alpha(v,b) - \lambda)(q^d + t \frac{\partial q^d}{\partial t}) \right] dF_b(b|v) dF_v(v)$$
$$= \int_v \int_b \left[(\lambda t - \alpha(v,b) b) \frac{\partial q^d}{\partial t} - (\alpha(v,b) - \lambda) q^d \right] dF_b(b|v) dF_v(v).$$

Using that t is constant, we can rewrite the equation as follows:

$$\frac{\partial V}{\partial t} = \lambda t \frac{\partial \bar{q}^d}{\partial t} - \int_v \int_b \left[\alpha(v,b) b \frac{\partial q^d}{\partial t} + (\alpha(v,b) - \lambda) q^d \right] dF_b(b|v) dF_v(v),$$

where the change in average demand $\bar{q}^d$ in response to a tax increase is given by $\frac{\partial \bar{q}^d}{\partial t} = \int_v \int_b \left[\frac{\partial q^d}{\partial t} \right] dF_b(b|v) dF_v(v)$. The optimal tax t^* is given by $\frac{\partial V}{\partial t}|_{t=t^*} \stackrel{!}{=} 0$, which gives:

$$\frac{\partial V}{\partial t} = t \frac{\partial \bar{q}^d}{\partial t} - \int_v \int_b \left[\hat{g}(v,b) b \frac{\partial q^d}{\partial t} - (1 - \hat{g}(v,b)) q^d \right] dF_b(b|v) dF_v(v) = 0$$

Abstracting from redistributive motives ($\hat{g}(v,b) = 1$ for all (v,b)) implies

$$t^* = \int_v \int_b b \left[\frac{\partial q^d}{\partial t} / \frac{\partial \bar{q}^d}{\partial t} \right] dF_b(b|v) dF_v(v), \tag{22}$$

where $\frac{\partial q^d}{\partial t} / \frac{\partial \bar{q}^d}{\partial t}$ denotes the relative responsiveness of a consumer type (v,b), i.e., the change in demand for that consumer type in response to a tax increase, relative to change in total demand.

If $c'''(\cdot) = 0$, the optimal linear tax simplifies to $t^* = E[b]$. To see this, differentiate Equation (11) with respect to t, which yields $\frac{\partial q^d}{\partial t} = -\frac{1}{c''(q^d)}$. This term is constant if $c'''(\cdot) = 0$, so that $\partial q^d / \partial t = \partial \bar{q}^d / \partial t$. Hence, the optimal linear tax is $t^* = E[b]$.

A.C Proof of Proposition 2: Condition for Implementability

Definition 2 (Internality Tax Implementability). *The allocation function* $\xi : \left[\underline{\hat{v}}, \overline{\hat{v}}\right] \to Q$ *is internality-tax implementable if there exists a tax function* $\tau : \left[\underline{\hat{v}}, \overline{\hat{v}}\right] \to \mathbb{R}$ *such that* $\left\{(\xi(\hat{v}), \tau(\hat{v}))|\hat{v} \in \left[\underline{\hat{v}}, \overline{\hat{v}}\right]\right\}$ *satisfy incentive compatibility according to*

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

Implementability hinges on two necessary conditions, which concern the *consumer preferences* or directly stem from them. First, the consumer's utility function must satisfy a single-crossing condition. Second, ξ must be monotonic in $\hat{v}$. Since, in our setting, the single-crossing condition is satisfied via $\partial\left(u_q^d/u_t^d\right)/\partial\hat{v} < 0$, a necessary condition for implementability is that the allocation is non-decreasing in perceived types, i.e., $\partial\xi/\partial\hat{v} \geq 0$ (Proof: See Appendix A.C.1).

Implementability of an incentive compatible mechanism additionally hinges on conditions, which stem from the *policy maker's preferences*. In our setting, these involve (paternalistic) corrective and redistributive motives. In a setting without corrective motives, a sufficient condition for this requirement involves the hazard rate of F, that is, $f(\hat{v})/(1 - F(\hat{v}))$, which is a measure of the thinness of the tail of the distribution. For our setting, Proposition 2 shows that the condition is more restrictive involving terms that stem from the corrective motive (Proof: See Appendix A.C.2).

A.C.1 Proof: Single-Crossing and Monotonicity

Incentive compatibility requires

$$\hat{v} = \arg\max_{\tilde{v}} u^d\left(\xi(\tilde{v}), \tau(\tilde{v}), \hat{v}\right) \quad \forall \tilde{v} \in \hat{V}$$

The first-order condition implies

$$\begin{aligned} &\left.\frac{\partial u^d\left(\xi(\tilde{v}), \tau(\tilde{v}), \hat{v}\right)}{\partial \tilde{v}}\right|_{\tilde{v}=\hat{v}} =: \left.\frac{\partial u^d\left(\tilde{v}, \hat{v}\right)}{\partial \tilde{v}}\right|_{\tilde{v}=\hat{v}} = 0 \\ \Leftrightarrow\; &\frac{\partial u^d\left(\tilde{v}, \hat{v}\right)}{\partial \xi} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} + \frac{\partial u^d\left(\tilde{v}, \hat{v}\right)}{\partial \tau} \cdot \frac{\partial \tau(\tilde{v})}{\partial \tilde{v}} = 0 \\ &\Leftrightarrow \frac{\partial \tau(\tilde{v})}{\partial \tilde{v}} = -\frac{\partial u^d\left(\tilde{v}, \hat{v}\right)/\partial \xi}{\partial u^d\left(\tilde{v}, \hat{v}\right)/\partial \tau} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}}. \end{aligned} \tag{23}$$

Differentiating the first-order condition with respect to $\hat{v}$ yields

$$\left.\frac{\partial^2 u^d\left(\tilde{v}, \hat{v}\right)}{\partial \tilde{v}^2}\right|_{\tilde{v}=\hat{v}} + \left.\frac{\partial^2 u^d\left(\tilde{v}, \hat{v}\right)}{\partial \tilde{v} \partial \hat{v}}\right|_{\tilde{v}=\hat{v}} = 0. \tag{24}$$

The second-order condition implies that

$$\left.\frac{\partial^2 u^d\left(\tilde{v},\hat{v}\right)}{\partial \tilde{v}^2}\right|_{\tilde{v}=\hat{v}} \leq 0. \tag{25}$$

Equations (25) and (24) imply

$$\begin{aligned}
&\left.\frac{\partial u^d\left(\tilde{v},\hat{v}\right)}{\partial \tilde{v}\partial \hat{v}}\right|_{\tilde{v}=\hat{v}} \geq 0 \\
\Leftrightarrow\ & \frac{\partial\left(\frac{\partial u^d(\tilde{v},\hat{v})}{\partial \xi}\right)}{\partial \hat{v}} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} + \frac{\partial\left(\frac{\partial u^d(\tilde{v},\hat{v})}{\partial \tau}\right)}{\partial \hat{v}} \cdot \frac{\partial \tau(\tilde{v})}{\partial \tilde{v}} \geq 0 \\
\overset{(23)}{\Leftrightarrow}\ & \frac{\partial\left(\frac{\partial u^d}{\partial \xi}\right)}{\partial \hat{v}} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} - \frac{\partial\left(\frac{\partial u^d}{\partial \tau}\right)}{\partial \hat{v}} \cdot \frac{\partial u^d/\partial \xi}{\partial u^d/\partial \tau} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} \geq 0 \\
\Leftrightarrow\ & \frac{\frac{\partial\left(\frac{\partial u^d}{\partial \xi}\right)}{\partial \hat{v}} \partial u^d/\partial \tau - \frac{\partial\left(\frac{\partial u^d}{\partial \tau}\right)}{\partial \hat{v}} \cdot \partial u^d/\partial \xi}{\partial u^d/\partial \tau} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} \geq 0 \\
\Leftrightarrow\ & \frac{\frac{\partial\left(\frac{\partial u^d}{\partial \xi}\right)}{\partial \hat{v}} \partial u^d/\partial \tau - \frac{\partial\left(\frac{\partial u^d}{\partial \tau}\right)}{\partial \hat{v}} \cdot \partial u^d/\partial \xi}{\left(\partial u^d/\partial \tau\right)^2} \cdot \frac{\partial u^d}{\partial \tau} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} \geq 0 \\
\Leftrightarrow\ & \frac{\partial\left(\frac{\partial u^d/\partial \xi}{\partial u^d/\partial \tau}\right)}{\partial \hat{v}} \cdot \frac{\partial u^d}{\partial \tau} \cdot \frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} \geq 0,
\end{aligned}$$

where the second-but-last step follows from the quotient rule. Since we know that $\frac{\partial u^d}{\partial \tau} < 0$ and via single crossing $\frac{\partial\left(\frac{\partial u^d/\partial \xi}{\partial u^d/\partial \tau}\right)}{\partial \hat{v}} = \frac{\partial\left(\frac{\hat{v}-c'-t'}{-1}\right)}{\partial \hat{v}} = -1 < 0$, it follows that we need monotonicity of the form $\frac{\partial \xi(\tilde{v})}{\partial \tilde{v}} \geq 0$.

A.C.2 Proof: Internality Tax Implementability

Inserting the optimal smooth tax from Proposition 1, the consumer's first-order condition can be rewritten as

$$c'[\xi(\hat{v})] \overset{!}{=} \underbrace{\hat{v} - \hat{g}(\hat{v})E[b|\hat{v}] - (1-\hat{G}(\hat{v}))\frac{1-F(\hat{v})}{f(\hat{v})}}_{=:\phi(\hat{v})}. \tag{26}$$

An incentive compatible mechanism must guarantee that ξ is non-decreasing in $\hat{v}$ (see Appendix A.C.1). Since c is convex, ξ is non-decreasing in $\hat{v}$ if and only if the left-hand side of Equation (26) is non-decreasing in $\hat{v}$. Accordingly, the right-hand side of Equation (26) must be non-decreasing in $\hat{v}$ as well, which implies

$$1 - \frac{\partial \hat{g}(\hat{v})}{\partial \hat{v}} E[b|\hat{v}] - \frac{\partial E[b|\hat{v}]}{\partial \hat{v}} \hat{g}(\hat{v}) - \frac{\partial[(1-\hat{G}(\hat{v}))\cdot(\{1-F(\hat{v})\}/f(\hat{v}))]}{\partial \hat{v}} \geq 0. \tag{27}$$

A.D Proof of Proposition 3: Bunching

In this section, we show that a failure of Internality Tax implementability leads to bunching and discuss how bunching at the top or at the bottom is equivalent to a ban for high or low attribute levels. To determine the optimal policy when bunching occurs, we apply the approach by Guesnerie and Laffont (1984) to conduct "ironing" using optimal control theory.

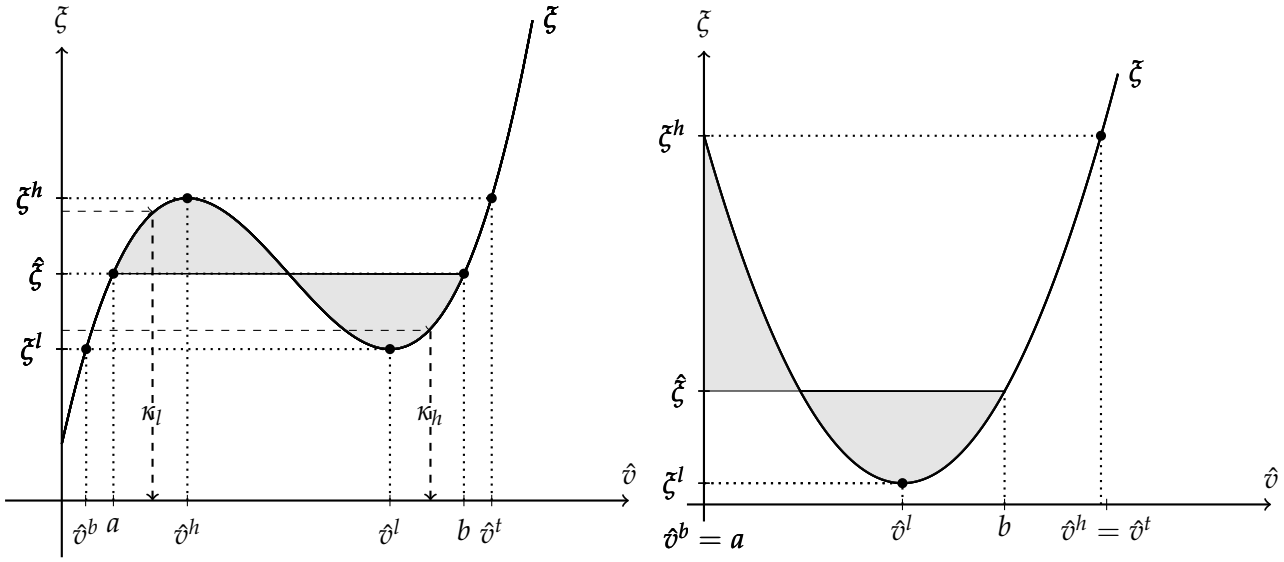
Before we present the formal method of ironing, we illustrate the intuition of this procedure and explain why it implies different forms of standards (uniform, maximum, and minimum). For the ease of exposition we abstract from redistributive motives ($g = 1 = G$). Suppose the smooth tax derived in Proposition 1 implies that the allocation rule ξ is decreasing in the interval $[\hat{v}^h, \hat{v}^l]$, where $\hat{v}^h$ denotes the level of perceived valuation that leads to a higher ξ, while $\hat{v}^l$ leads to a lower ξ ($\hat{v}^h < \hat{v}^l$, but $\xi(\hat{v}^h) > \xi(\hat{v}^l)$). The goal of ironing is to determine over which interval $[a, b]$ of the type space bunching will take place.

Recall that an allocation rule that decreases in $\hat{v}$ implies a fundamental misalignment of preferences violating the condition in Proposition 2: the policy maker would like to allocate a lower allocation level to a consumer with a higher perceived valuation. However, incentive compatibility requires a non-decreasing allocation rule (see Appendix A.C.1): a consumer will not reveal a higher perceived valuation, if he obtains a lower attribute level than a consumer with a lower perceived valuation.

To ensure incentive compatibility, the policy maker resorts to bunching and assigns the same product attribute $\xi^* = \hat{\xi}$ to a subset of consumers. Suppose to the contrary that the implemented allocation rule ξ were partly increasing in the interval $[\hat{v}^h, \hat{v}^l]$. Then, the welfare of types that underconsume compared to ξ can be increased by increasing the allocation, while the welfare of types that overconsume can be increased by decreasing the allocation. This is true up to the point where the consumption of all types $\hat{v} \in [a, b]$ is equal to $\hat{\xi}$.

The standard $\hat{\xi}$ is set such that it minimizes the loss in welfare compared to the solution candidate ξ. Intuitively, the standard $\hat{\xi}$ lies in between the extremes of ξ^l and ξ^h so that some consumers overconsume compared to candidate ξ and some underconsume. If all consumers underconsume, i.e., $\hat{\xi} = \xi^l$, then increasing the standard marginally would raise the welfare of all consumers that still underconsume after the increase, while it would only decrease the welfare of the consumers who overconsume after the increase, that is, for types with an ideal consumption $\xi = \xi^l$. An analogous statement holds for the case when all consumers overconsume, i.e., $\hat{\xi} = \xi^h$.

The exact procedure of determining the types that over- or underconsume compared to ξ is described below for the general case of interior bunching, which sub-



(a) Interior bunching

(b) Bunching at the bottom: Banning low attribute levels

Figure 1: Ironing yields the optimal bunching regions

sumes all cases relevant to our scenario (uniform, minimum and maximum standards). Figure 1 b) visualizes the specific case of minimum standards, i.e., bunching at the bottom.

We now present the formal method of ironing. Bunching occurs if ξ (or equivalently, the term $\phi(\hat{v})$ as defined in Equation (26)) is decreasing over an interval $[\hat{v}^h, \hat{v}^l] \subseteq [\underline{\hat{v}}, \overline{\hat{v}}]$. Let ξ denote the non-ironed "solution candidate" and let ξ^* denote the correct solution involving ironing. Since we cannot guarantee that the allocation is strictly increasing, write

$$\nu(\hat{v}) := \tfrac{\partial \xi}{\partial \hat{v}} \geq 0. \tag{28}$$

We now have a control problem with an inequality constraint on the control ν. The Hamiltonian $\mathfrak{H}(\hat{v}, \xi, \nu, \delta)$ of the general optimization problem without implicitly assuming that the incentive constraints hold, can be written as

$$\mathfrak{H}(\cdot) = \left[\alpha(\hat{v}) \cdot \underbrace{\left(\hat{u}^d(\hat{v}) - E[b|\hat{v}]\xi(\hat{v}) \right)}_{=E[\hat{u}^n(\hat{v},b)|\hat{v}]} + \lambda \underbrace{\left(m + \hat{v}\xi(\hat{v}) - \hat{u}^d(\hat{v}) - c(\xi(\hat{v})) \right)}_{=\tau(\hat{v})} \right] f(\hat{v}) + \delta(\hat{v})\nu(\hat{v}),$$

where δ denotes the multiplier for the constraint given by Inequality (28). The first-order condition on the state ξ yields

$$\frac{-\delta'(\hat{v})}{\lambda} \overset{!}{=} \left[\hat{v} - \hat{g}(\hat{v}) E[b|\hat{v}] - c'(\xi^*(\hat{v})) \right] f(\hat{v}). \tag{29}$$

Without redistributive motives ($\hat{G} \equiv 1$), the term $\phi(\hat{v})$ as defined in Equation (26) is given by $\phi(\hat{v}) = \hat{v} - \hat{g}(\hat{v})E[b|\hat{v}]$. For the ease of exposition, we discuss the case without redistributive motives. Thus, Equation (29) can be rewritten as

$$\frac{-\delta'(\hat{v})}{\lambda} \stackrel{!}{=} \left[\phi(\hat{v}) - c'(\xi^*(\hat{v}))\right] f(\hat{v}). \tag{30}$$

For simplicity, let us assume there is one unique interval $[a, b]$ of bunching.[21] Then, integrating (30) between a and b, and using the transversality conditions $\delta(a) = \delta(b) = 0$, which hold since the monotonicity constraint is non-binding at the boundaries, yields

$$\int_a^b \left[\phi(\theta) - c'(\xi^*(\theta))\right] f(\theta)d\theta \stackrel{!}{=} \frac{1}{\lambda}\int_a^b -\delta'(\theta)f(\theta)d\theta = \frac{-\delta(b) + \delta(a)}{\lambda} = 0, \tag{31}$$

which implies that the average difference between the ϕ and the marginal costs (i.e., the average distortion of the "virtual surplus") is zero over the bunching interval.[22] Equation (31) together with the fact that $\xi^*(a) = \xi^*(b) = \xi(a) = \xi(b)$ characterizes the allocation $\xi^*(\hat{v}) = \hat{\xi}$ given to types $\hat{v} \in [a, b]$ in the optimal mechanism.

We now use the above characterization of the bunching region to determine its boundaries a and b and the allocation $\hat{\xi}$. The procedure is illustrated in Figure 1. We first consider a case where the bunching region is in the middle of the range of $\hat{v}$. If $a > \underline{v}$ and $b < \overline{v}$, then there exists $\hat{v}^l := \arg\min_{\hat{v}} \xi(\hat{v})$ s.t. $\hat{v} \in [a, b]$ and $\hat{v}^h := \arg\max_{\hat{v}} \xi(\hat{v})$ s.t. $\hat{v} \in [a, b]$. Denote $\xi^h := \xi(\hat{v}^h)$ and $\xi^l := \xi(\hat{v}^l)$. Let κ_l denote the inverse function of the increasing part of ξ defined on $[\hat{v}^b, \hat{v}^h]$ with $\hat{v}^b := \min_{\hat{v}} \hat{v}$ s.t. $\xi(\hat{v}) = \xi^l$. Analogously, κ_h denotes the inverse function of the increasing part of ξ defined on $[\hat{v}^l, \hat{v}^t]$ with $\hat{v}^t := \max_{\hat{v}} \hat{v}$ s.t. $\xi(\hat{v}) = \xi^h$. For $\tilde{\xi} \in [\xi^h, \xi^l]$, define

$$\Delta(\tilde{\xi}) := \int_{\kappa_l(\tilde{\xi})}^{\kappa_h(\tilde{\xi})} \left[\phi(\theta) - c'(\tilde{\xi})\right] f(\theta)d\theta. \tag{32}$$

Since $\xi^h > \xi(\hat{v})$, $\forall \hat{v} \in (\hat{v}^h, \hat{v}^t)$, it holds that $\Delta(\xi^h) < 0$. Analogously, since $\xi^l < \xi(\hat{v})$, $\forall \hat{v} \in (\hat{v}^b, \hat{v}^l)$, it holds that $\Delta(\xi^l) > 0$. Therefore, by the intermediate value theorem, there must exists some $\hat{\xi}$, such that $\Delta(\hat{\xi}) = 0$ as required by Equation (31). Thus, $a = \kappa_l(\hat{\xi})$ and $b = \kappa_h(\hat{\xi})$. As a consequence, all individuals with $\hat{v} \in [a, b]$ will be assigned the same level of the product attribute $\hat{\xi}$.

[21] The extension to a setting with several bunching regions is only subject to minor caveats. See Guesnerie and Laffont (1984) for details.

[22] Note that in a mechanism which optimally does not involve bunching, the difference between ϕ and the marginal costs is zero at each point as can be seen in Equation (26).

If $a = \underline{\hat{v}}$ (analogous reasoning applies for $b = \overline{\hat{v}}$), bunching occurs at the bottom of the type distribution (see Panel b of Figure 1). In that case, we need to define and evaluate

$$\Delta(\tilde{\xi}) := \int_{\underline{\hat{v}}}^{\kappa_h(\tilde{\xi})} \left[\phi(\theta) - c'(\tilde{\xi})\right] f(\theta) d\theta. \tag{33}$$

As a consequence, all individuals with $\hat{v} \in [\underline{\hat{v}}, b]$ will be assigned the same level of the product attribute $\hat{\xi}$. As bunching occurs at the bottom, this outcome corresponds to a minimum standard, i.e., a ban of all realizations of the product attribute below $\hat{\xi}$. The rationale is equivalent for bunching at the top.

A.E Proof of Equation (5): First-Order Approximation

As a consequence of the Regression Conditional Expectation Function Theorem (e.g., Angrist and Pischke, 2009), the Minimum Mean Squared Error (MMSE) linear approximation of the conditional expectation $E\left[b|\hat{v}\right]$ is given by:

$$\hat{E}\left[b|\hat{v}\right] = E(b|\hat{v} = \mu_{\hat{v}}) + \frac{cov(b,\hat{v})}{\sigma^2_{\hat{v}}} \cdot \left[\hat{v} - \mu_{\hat{v}}\right]. \tag{34}$$

Using that $cov(b,\hat{v}) = cov(b,v) + \sigma^2_b$, $\sigma^2_{\hat{v}} = \sigma^2_v + \sigma^2_b + 2\,cov(b,v)$, and $cov(b,v) = \rho\sigma_v\sigma_b$, we obtain after rearranging:

$$\hat{E}\left[b|\hat{v}\right] = E\left[b|\mu_{\hat{v}}\right] + \frac{\rho + (\sigma_b/\sigma_v)}{(\sigma_b/\sigma_v) + (\sigma_v/\sigma_b) + 2\rho} \cdot \left[\hat{v} - \mu_{\hat{v}}\right]. \tag{35}$$

The existence of the conditional expectation as a function of $\hat{v}$ is guaranteed by the Factorization Lemma and the Radon-Nikodym theorem.

In the bivariate normal case, (v,b) is jointly normal distributed with:

$$(v,b) \sim N\left(\left[\begin{array}{c} \mu_v \\ \mu_b \end{array}\right], \left[\begin{array}{cc} \sigma^2_v & \rho\sigma_v\sigma_b \\ \rho\sigma_v\sigma_b & \sigma^2_b \end{array}\right]\right). \tag{36}$$

Hence, the perceived valuation $\hat{v} = v + b$ has the following normal distribution

$$\hat{v} \sim N\left(\mu_v + \mu_b, \sigma^2_v + \sigma^2_b + 2\rho\sigma_v\sigma_b\right) =: \left(\mu_{\hat{v}}, \sigma^2_{\hat{v}}\right). \tag{37}$$

The conditional expectation $E[b|\hat{v}]$ of the bias can be calculated as

$$E\left[b|\hat{v}\right] = \frac{cov(b,\hat{v})}{\sigma^2_{\hat{v}}} \cdot \hat{v} + \left[\mu_b - \frac{cov(b,\hat{v})}{\sigma^2_{\hat{v}}} \cdot \mu_{\hat{v}}\right] = \mu_b + \frac{cov(b,\hat{v})}{\sigma^2_{\hat{v}}} \cdot \left[\hat{v} - \mu_{\hat{v}}\right], \tag{38}$$

with $cov(b, \hat{v}) = cov(b, v) + \sigma_b^2$. Rearranging gives that

$$E\left[b|\hat{v}\right] = \mu_b + \frac{\rho + (\sigma_b/\sigma_v)}{(\sigma_b/\sigma_v) + (\sigma_v/\sigma_b) + 2\rho} \cdot \left[\hat{v} - \mu_{\hat{v}}\right]. \tag{39}$$

Hence, the approximation error $R(\hat{v})$ equals zero for all $\hat{v}$ if v and b follow a bivariate normal distribution.

A.F Proof of Proposition 6: Prices Vs. Quantities

The proof proceeds in two steps. In a first step, we explore the welfare gains from non-linear taxation relative to linear taxes and to standards. The corresponding results are summarized in Proposition 8. In a second step, we then use these results to compare linear taxes with standards. We analyze a setting in which the local bias heterogeneity A is constant.

Proposition 8 (Welfare Gain from Non-Linear Taxation)**.** *The expected welfare gain of the optimal non-linear tax relative to the optimal linear tax and the optimal standard is weakly positive. It depends on the local bias heterogeneity A as follows:*

a) The expected welfare gain over the optimal linear tax is zero when $A = 0$ and increases in the absolute value of A;

b) The expected welfare gain over the optimal standard is zero when $A \geq 1$ and decreases in A.

We now provide the proof of Proposition 8. We analyze the impact of a change in the local bias heterogeneity A on expected welfare under the optimal non-linear tax, the optimal linear tax, and the optimal standard. To separate the impact of changes in A from changes in the population of consumers, we hold the distribution of $\hat{v}$ as well as the first moments of v and b constant. In addition, we consider a scenario in which redistribution does not matter, i.e., $\hat{g}(\hat{v}) = 1$.

To evaluate welfare implications of a change in A, we need to evaluate the derivative with respect to A of the expected equilibrium normative utility (net of taxes) for consumers with $\hat{v}$ (see also the Hamiltonian of the policy maker's problem):

$$\int_{\hat{v}} \hat{u}^n(A) dF(\hat{v}) = \int_{\hat{v}} E[v|\hat{v}](A) q^M(A) - c\left(q^M(A)\right) dF(\hat{v}), \tag{40}$$

where q^M denotes the allocation under the non-linear, **p**rice (linear tax) or **s**tandard (consumption ban) mechanism, $q^M(A) \in \{\xi(A), q^P(A), q^S\}$, and $\hat{u}^n(A, q^M)$ denotes the equilibrium normative utility in the respective mechanism.

We first consider the welfare implications of a change in A under the optimal non-linear tax. We have that:

$$\frac{\partial \hat{u}^n(A,\xi)}{\partial A} = \frac{\partial E[v|\hat{v}]}{\partial A}\xi + \frac{\partial \xi}{\partial A}\{\underbrace{E[v|\hat{v}] - c'(\xi)}_{=0}\}.$$

By an envelope theorem argument, the second summand is zero since $E[v|\hat{v}] = \hat{v} - E[b|\hat{v}] = c'(\xi)$ due to optimal consumer behavior (see Equation (11)).

The optimal standard q^S can be calculated by solving the following utility maximization problem: $\max_{q\in\mathbb{R}} \int_v u^n(q,v)dF_v(v)$. Inserting $u^n(q,v) = m + vx - c(q)$ and solving for the maximum yields the following implicit solution: $c'(q^S) = E[v]$. Hence, q^S does not depend on A and we obtain:

$$\frac{\partial \hat{u}^n(A,q^S)}{\partial A} = \frac{\partial E[v|\hat{v}]}{\partial A}q^S.$$

As shown in Section A.B, the optimal linear tax is given by the following expression: $t^*(A) = \int_{\hat{v}} E(b|\hat{v})(A)\left[\frac{\partial q^d}{\partial t} / \frac{\partial \bar{q}^d}{\partial t}\right] dF(\hat{v})$. From the first order condition of consumer maximization, we know that $c'(q^d) = \hat{v} - t^*(A) = E[v|\hat{v}] + E[b|\hat{v}] - t^*(A)$. Using this equation, we obtain:

$$\frac{\partial \hat{u}^n(A,q^P)}{\partial A} = \frac{\partial E[v|\hat{v}]}{\partial A}q^P + \frac{\partial q^P}{\partial A}\{E[v|\hat{v}] - c'(q^P)\} = \frac{\partial E[v|\hat{v}]}{\partial A}q^P + \frac{\partial q^P}{\partial A}\{E[b|\hat{v}] - t^*(A)\}.$$

The second summand vanishes when taking the integral over all types, which can be seen as follows:

$$\begin{aligned}&\int_{\hat{v}} \frac{\partial q^P}{\partial A}\{E[b|\hat{v}] - t^*(A)\}dF(\hat{v}) = \frac{\partial t^*(A)}{\partial A}\int_{\hat{v}} \frac{\partial q^P}{\partial t^*}\{E[b|\hat{v}] - t^*(A)\}dF(\hat{v})\\ =&\frac{\partial t^*(A)}{\partial A}\left(\int_{\hat{v}} \frac{\partial q^P}{\partial t^*}E[b|\hat{v}]dF(\hat{v}) - t^*(A)\int_{\hat{v}} \frac{\partial q^P}{\partial t^*}dF(\hat{v})\right) = 0,\end{aligned}$$

where the last equality follows from the definition of $t^*(A)$ and the fact that $\frac{\partial \bar{q}^d}{\partial t^*} = \int_{\hat{v}} \frac{\partial q^P}{\partial t^*}dF(\hat{v})$. Hence, we obtain:

$$\frac{\partial \hat{u}^n(A,q^P)}{\partial A} = \frac{\partial E[v|\hat{v}]}{\partial A}q^P.$$

We now evaluate how a change in A changes the relative advantage of non-linear taxation relative to a standard in terms of expected welfare:

$$\frac{\partial \Delta u^n_{NL,S}}{\partial A} = \int \frac{\partial E[v|\hat{v}]}{\partial A}\left(\xi - q^S\right) dF(\hat{v}),$$

where $\Delta u^n_{NL,S} := \int \hat{u}^n(A,\xi) - \hat{u}^n(A,q^S)dF(\hat{v})$. Using $\frac{\partial E[v|\hat{v}]}{\partial A} = (\mu_{\hat{v}} - \hat{v})$, we obtain:

$$\begin{aligned}
\frac{\partial \Delta u^n_{NL,S}}{\partial A} &= \int (\mu_{\hat{v}} - \hat{v}) \left[\xi(\hat{v}) - q^S\right] dF(\hat{v}) \\
&= \mu_{\hat{v}} \int \xi(\hat{v}) dF(\hat{v}) - \int \hat{v}\xi(\hat{v}) dF(\hat{v}) - q^S \left[\mu_{\hat{v}} - \int \hat{v} dF(\hat{v})\right] \\
&= E[\hat{v}] \cdot E\left[\xi(\hat{v})\right] - E\left[\hat{v} \cdot \xi(\hat{v})\right] - 0 \\
&= E[\hat{v}] \cdot E\left[\xi(\hat{v})\right] - E[\hat{v}] \cdot E\left[\xi(\hat{v})\right] - cov\left(\hat{v}, \xi(\hat{v})\right) \leq 0,
\end{aligned}$$

where the last inequality holds since ξ is increasing in $\hat{v}$ for an incentive compatible mechanism, so that $cov\left(\hat{v}, \xi(\hat{v})\right)$ is positive.

The relative advantage in terms of expected welfare of non-linear taxation relative to the optimal linear tax is given by:

$$\frac{\partial \Delta u^n_{NL,P}}{\partial A} = \int \frac{\partial E[v|\hat{v}]}{\partial A} \left(\xi(\hat{v}) - q^P(\hat{v})\right) dF\hat{v}.$$

We approximate $\xi - q^P$ by a first-order Taylor approximation in the marginal tax rate t' around $t' = t^*$, which yields:

$$\xi - q^P \approx \left.\frac{dq^P}{dt}\right|_{t'=t^*} \left[t'(\hat{v}) - t^*\right],$$

where $t'(\hat{v}) = E[b|\hat{v}]$ is the optimal non-linear tax rate from Proposition 1 and t^* is the optimal linear tax rate from Equation (22).

Using that $E[b|\hat{v}] = E[b|\mu_{\hat{v}}] + A(\hat{v} - \mu_{\hat{v}})$, rearranging, and using $W(\hat{v}) := \frac{dq(\hat{v})/dt}{d\bar{q}/dt}$ with $\frac{d\bar{q}}{dt} := \int_{\hat{v}} \frac{dq(\hat{v})}{dt} dF(\hat{v})$ to denote the relative responsiveness of a consumer type $\hat{v}$ we obtain:

$$\xi - q^P \approx \left.\frac{dq^P}{dt}\right|_{t'=t^*} A\left(\hat{v} - \int \hat{v} W(\hat{v}) dF(\hat{v})\right).$$

Denoting $v^P = \int \hat{v} W(\hat{v}) dF(\hat{v})$, we have:

$$\begin{aligned}
\frac{\partial \Delta u^e_{NL,P}}{\partial A} &= A \int (\mu_{\hat{v}} - \hat{v})(\hat{v} - v^P) \left.\frac{dq^P}{dt}\right|_{t'=t^*} dF(\hat{v}) \\
&= A \int (\mu_{\hat{v}}\hat{v} - \mu_{\hat{v}} v^P - \hat{v}^2 + \hat{v} v^P) \left.\frac{dq^P}{dt}\right|_{t'=t^*} dF(\hat{v}) \\
&= A \left((\mu_{\hat{v}} + v^P) \int \hat{v} \left.\frac{dq^P}{dt}\right|_{t'=t^*} dF(\hat{v}) - (\mu_{\hat{v}} v^P) \int \left.\frac{dq^P}{dt}\right|_{t'=t^*} dF(\hat{v}) - \int \hat{v}^2 \left.\frac{dq^P}{dt}\right|_{t'=t^*} dF(\hat{v}) \right) \\
&= A \frac{d\bar{q}^P}{dt} \left((\mu_{\hat{v}} + v^P) \int \hat{v} W(\hat{v}) dF(\hat{v}) - \mu_{\hat{v}} v^P - \int \hat{v}^2 W(\hat{v}) dF(\hat{v}) \right)
\end{aligned}$$

$$= A\frac{d\bar{q}^P}{dt}\left((v^P)^2 - \int \hat{v}^2 W(\hat{v}) dF(\hat{v})\right)$$
$$= \underbrace{A\frac{d\bar{q}^P}{dt}}_{\leq 0 \text{ if } A \geq 0} \underbrace{\left(\left(\int \hat{v} W(\hat{v}) dF(\hat{v})\right)^2 - \int \hat{v}^2 W(\hat{v}) dF(\hat{v})\right)}_{\leq 0} \geq 0,$$

where $\frac{d\bar{q}^P}{dt} = \int \frac{dq^P}{dt}\big|_{t'=t^*} dF(\hat{v})$. The fact that the third factor is non-positive follows from Jensen's inequality because $d(\hat{v}) = W(\hat{v}) f(\hat{v})$ satisfies the properties of a density function, with $d(\hat{v}) \geq 0 \; \forall \, \hat{v}$ and $\int d(\hat{v}) d\hat{v} = \int W(\hat{v}) dF(\hat{v}) = 1$.

Comparison of Linear Price Instruments with Standards

In Proposition 8, we have shown that a linear price instrument is welfare-optimal for $A = 0$, a standard is optimal for $A = 1$, and the advantage of a standard compared to a linear price instrument increases in A. Hence, the existence of $\hat{A}$ from Proposition 6 is guaranteed by applying the intermediate value theorem.

To complete the proof of Proposition 6, we need to show that the welfare advantage of a price instrument over a quantity instrument becomes more pronounced as A decreases below zero. For that purpose, it is sufficient to show that the difference in the expected equilibrium normative utility between the linear tax and the quantity regulation, $\partial \Delta u^n_{P,S} / \partial A = \int \hat{u}^n(A, q^P) - \hat{u}^n(A, q^S) dF(\hat{v})$, decreases in A. Let $\bar{v}$ be the $\hat{v}$ such that $q^P(\hat{v}) = q^S$. We approximate the demand function by a first-order Taylor approximation at $\bar{v}$, which yields $q^P(\hat{v}) \approx q^P(\bar{v}) + \frac{\partial q^P}{\partial \hat{v}}\big|_{\bar{v}} (\hat{v} - \bar{v})$. As the first-order conditions of utility maximization imply $\hat{v} - t^* = c'\left(q^P(\hat{v})\right)$ for linear taxation and $c'\left(q^S\right) = E[v]$ for quantity regulation, we have that $\bar{v} = E[v] + t^*$, where $t^* = \int_{\hat{v}} E[b|\hat{v}] \, W(\hat{v}) \, dF(\hat{v})$. We can now show that the welfare difference between prices and quantities decreases in A, since

$$\begin{aligned}
\frac{\partial \Delta u^n_{P,S}}{\partial A} &= \int_{\hat{v}} \frac{\partial E[v|\hat{v}]}{\partial A} \left[q^P(\hat{v}) - q^S\right] dF(\hat{v}) \\
&= \int_{\hat{v}} (\mu_{\hat{v}} - \hat{v}) \left[q^P(\hat{v}) - q^S\right] dF(\hat{v}) \\
&= -\frac{\partial q^P}{\partial \hat{v}}\big|_{\hat{v}=\bar{v}} \int_{\hat{v}} (\mu_{\hat{v}} - \hat{v})(\hat{v} - \bar{v}) \, dF(\hat{v}) \\
&= -\frac{\partial q^P}{\partial \hat{v}}\big|_{\hat{v}=\bar{v}} \int_{\hat{v}} (\mu_{\hat{v}} - \hat{v}) \left[(v - E[v])(b - t^*)\right] dF(\hat{v}) \\
&= \underbrace{-\frac{\partial q^P}{\partial \hat{v}}\big|_{\hat{v}=\bar{v}}}_{<0} \underbrace{\left[E\left[\hat{v}^2\right] - \mu^2_{\hat{v}}\right]}_{>0} < 0,
\end{aligned}$$

where the evaluation of the second factor in the last line follows from Jensen's inequality.

A.G Proof of Proposition 7: Minimum and Maximum Standards

The proof proceeds as follows. From Proposition 3, we know that $\partial E[v|\hat{v}]/\partial\hat{v} < 0$ for some $\hat{v}$ implies that a policy maker will use bunching as part of the optimal policy. Furthermore, we know from Appendix A.D that bans are part of the optimal policy mix if bunching occurs at the top or the bottom of the perceived valuation distribution. We use a second-order Taylor approximation to approximate $E[v|\hat{v}]$ as a function of $\hat{v}$ and $\hat{v}^2$. This allows to derive the conditions under which mixed policies optimally involve bans for high or low levels of the product attribute in terms of higher-order moments of the (joint) distributions of the random variables v, b, and $\hat{v}$.

The second-order Taylor approximation of $E\left[v|\hat{v}\right]$ is:

$$\hat{E}\left[v|\hat{v}\right] = c_0 + c_1\hat{v} + c_2\hat{v}^2. \tag{41}$$

Bunching occurs whenever $\partial E\left[v|\hat{v}\right]/\partial\hat{v} = c_1 + 2c_2\hat{v} \leq 0$ for some $\hat{v}$. For the purpose of this proof, we are interested in bunching at the top and at the bottom. With $\hat{v} \in [0, \infty)$, we have:

1. Bunching at the Top if $c_1 > 0$ and $c_2 < 0$.
2. Bunching at the Bottom if $c_1 < 0$ and $c_2 > 0$.

Otherwise, bunching takes place over the entire support of $\hat{v}$ ($c_1 < 0, c_2 \leq 0$) or not at all ($c_1 > 0, c_2 \geq 0$). We proceed by first calculating c_1 and c_2. In a subsequent step, we derive sufficient conditions for bunching at the top and at the bottom.

The Frisch-Waugh-Lovell Theorem demonstrates that c_r for $r \in \{1, 2\}$, as defined by Equation (41), can be calculated as

$$c_r = \frac{cov\left(\epsilon_r, \hat{\epsilon}_r\right)}{var\left(\hat{\epsilon}_r\right)}, \tag{42}$$

where $\hat{\epsilon}_r$ is the "residual" of $\hat{v}^r$ after regressing it on all covariates except $\hat{v}^r$, and ϵ_r is the "residual" of v after regressing it on all covariates except $\hat{v}^r$.

We first calculate c_1. Regressing $\hat{v}$ on $\hat{v}^2$ and a constant yields:

$$\begin{aligned}\hat{v} &= d_1 + d_2\hat{v}^2 + \hat{\epsilon}_1 \\ &= E\left[\hat{v}|\hat{v}^2 = 0\right] + \frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 + \hat{\epsilon}_1.\end{aligned}$$

The residual $\hat{\epsilon}_1$ is then given by:

$$\hat{\epsilon}_1 = \hat{v} - \left\{ E\left[\hat{v}|\hat{v}^2 = 0\right] + \frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 \right\}.$$

Regressing v on $\hat{v}^2$ and a constant yields:

$$\begin{aligned} v &= e_0 + e_1\hat{v}^2 + \epsilon_1 \\ &= E\left[v|\hat{v}^2 = 0\right] + \frac{cov\left(v, \hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 + \epsilon_1 \end{aligned}$$

The residual ϵ_1 is then given by:

$$\epsilon_1 = v - \left\{ E\left[v|\hat{v}^2 = 0\right] + \frac{cov\left(v, \hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 \right\}.$$

Therefore, by Equation (42), the coefficient c_1 is given by:

$$\begin{aligned} c_1 &= \frac{cov\left(\epsilon_1, \hat{\epsilon}_1\right)}{var\left(\hat{\epsilon}_1\right)} \\ &= \frac{cov\left(v - \left\{ E\left[v|\hat{v}^2 = 0\right] + \frac{cov\left(v,\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 \right\}, \hat{v} - \left\{ E\left[\hat{v}|\hat{v}^2 = 0\right] + \frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2 \right\}\right)}{var\left(\hat{\epsilon}_1\right)} \\ &= \frac{cov\left(v, \hat{v} - \frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} + \frac{cov\left(-\frac{cov\left(v,\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2, \hat{v} - \frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} \\ &= \frac{cov\left(v, \hat{v} - \frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} - \frac{\frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}cov\left(v, \hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} + \frac{\frac{cov\left(\hat{v},\hat{v}^2\right)cov\left(v,\hat{v}^2\right)}{\left[var\left(\hat{v}^2\right)\right]^2}cov\left(\hat{v}^2, \hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} \\ &= \frac{cov\left(v, \hat{v} - \frac{cov\left(\hat{v},\hat{v}^2\right)}{var\left(\hat{v}^2\right)}\hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)}. \end{aligned} \tag{43}$$

Next, we calculate c_2. Regressing $\hat{v}^2$ on $\hat{v}$ and a constant yields:

$$\begin{aligned} \hat{v}^2 &= d_0 + d_1\hat{v} + \hat{\epsilon}_2 \\ &= E\left[\hat{v}^2|\hat{v} = 0\right] + \frac{cov\left(\hat{v}^2, \hat{v}\right)}{var\left(\hat{v}\right)}\hat{v} + \hat{\epsilon}_2. \end{aligned}$$

The residual $\hat{\epsilon}_2$ is then given by:

$$\hat{\epsilon}_2 = \hat{v}^2 - \left\{ E\left[\hat{v}^2 | \hat{v} = 0\right] + \frac{cov\left(\hat{v}^2, \hat{v}\right)}{var\left(\hat{v}\right)} \hat{v} \right\}.$$

Regressing v on $\hat{v}$ and a constant yields:

$$\begin{aligned} v &= d_0 + d_1 \hat{v} + \epsilon_2 \\ &= E\left[v | \hat{v} = 0\right] + \frac{cov\left(v, \hat{v}\right)}{var\left(\hat{v}\right)} \hat{v} + \epsilon_2. \end{aligned}$$

The residual ϵ_2 is then given by:

$$\epsilon_2 = v - \left\{ E\left[v | \hat{v} = 0\right] + \frac{cov\left(v, \hat{v}\right)}{var\left(\hat{v}\right)} \hat{v} \right\}.$$

Therefore, by Equation (42), the coefficient c_2 is given by:

$$\begin{aligned} c_2 &= \frac{cov\left(\epsilon_2, \hat{\epsilon}_2\right)}{var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(v - \left\{ E\left[v | \hat{v} = 0\right] + \frac{cov(v,\hat{v})}{var(\hat{v})} \hat{v} \right\}, \hat{v}^2 - \left\{ E\left[\hat{v}^2 | \hat{v} = 0\right] + \frac{cov\left(\hat{v}^2,\hat{v}\right)}{var(\hat{v})} \hat{v} \right\}\right)}{var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(v, \hat{v}^2\right) - \frac{cov(v,\hat{v})}{var(\hat{v})} cov\left(\hat{v}^2, \hat{v}\right)}{var\left(\hat{\epsilon}_2\right)} + \frac{cov\left(-\frac{cov(v,\hat{v})}{var(\hat{v})}\hat{v}, \hat{v}^2 - \frac{cov\left(\hat{v}^2,\hat{v}\right)}{var(\hat{v})}\hat{v}\right)}{var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(v, \hat{v}^2\right) - \frac{cov(v,\hat{v})}{var(\hat{v})} \left[cov\left(\hat{v}^2, v\right) + cov\left(\hat{v}^2, b\right)\right]}{var\left(\hat{\epsilon}_2\right)} + \frac{-\frac{cov(v,\hat{v})}{var(\hat{v})} cov\left(\hat{v}, \hat{v}^2\right) + cov\left(v, \hat{v}\right) \frac{cov\left(\hat{v}^2,\hat{v}\right)}{var(\hat{v})}}{var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(v, \hat{v}^2\right) - \frac{cov(v,\hat{v})}{var(\hat{v})} \left[cov\left(\hat{v}^2, v\right) + cov\left(\hat{v}^2, b\right)\right]}{var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(v, \hat{v}^2\right) cov\left(b, \hat{v}\right) - cov\left(v, \hat{v}\right) cov\left(\hat{v}^2, b\right)}{var\left(\hat{v}\right) var\left(\hat{\epsilon}_2\right)} \\ &= \frac{cov\left(\hat{v}, \hat{v}^2\right) cov\left(b, \hat{v}\right) - cov\left(\hat{v}, \hat{v}\right) cov\left(\hat{v}^2, b\right)}{var\left(\hat{v}\right) var\left(\hat{\epsilon}_2\right)} \\ &\overset{(*)}{=} \frac{\left[2\mu_{\hat{v}}\sigma_{\hat{v}}^2 + sk(\hat{v})\sigma_{\hat{v}}^3\right] cov\left(b, \hat{v}\right) - \sigma_{\hat{v}}^2 cov\left(\hat{v}^2, b\right)}{var\left(\hat{v}\right) var\left(\hat{\epsilon}_2\right)} \\ &= \frac{\sigma_{\hat{v}}^2 \overbrace{\left\{\left[2\mu_{\hat{v}} + sk(\hat{v})\sigma_{\hat{v}}\right] cov\left(b, \hat{v}\right) - cov\left(\hat{v}^2, b\right)\right\}}^{=:K}}{var\left(\hat{v}\right) var\left(\hat{\epsilon}_2\right)}, \end{aligned} \tag{44}$$

where the denominator is always positive and $(*)$ follows from the fact that, for every random variable Y, we have:[23]

$$cov\left(Y, Y^2\right) = \left[sk(Y)\sigma_Y + 2\mu_Y\right]\sigma_Y^2. \tag{45}$$

We now analyze the term K:

$$\begin{aligned}
K &\overset{\text{BG1969}}{=} \left[2\mu_{\hat{v}} + sk(\hat{v})\sigma_{\hat{v}}\right] cov\left(b, \hat{v}\right) - \left[2\mu_{\hat{v}} cov\left(\hat{v}, b\right) + E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right]\right] \\
&= sk(\hat{v})\sigma_{\hat{v}} cov\left(b, \hat{v}\right) - E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right] \\
&= E\left[\left(\frac{\hat{v} - \mu_{\hat{v}}}{\sigma_{\hat{v}}}\right)^3\right] \sigma_{\hat{v}} E\left[\left(\hat{v} - \mu_{\hat{v}}\right)\left(b - \mu_b\right)\right] - E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right] \\
&= \frac{E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^3\right] E\left[\left(\hat{v} - \mu_{\hat{v}}\right)\left(b - \mu_b\right)\right]}{\sigma_{\hat{v}}^2} - E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right] \\
&= \frac{E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^3\right] E\left[\left(\hat{v} - \mu_{\hat{v}}\right)\left(b - \mu_b\right)\right]}{\sigma_{\hat{v}}^2} - E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right] \\
&= \sigma_{\hat{v}}^2 \sigma_b \left\{ \frac{E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^3\right] E\left[\left(\hat{v} - \mu_{\hat{v}}\right)\left(b - \mu_b\right)\right]}{\sigma_{\hat{v}}^4 \sigma_b} - \frac{E\left[\left(\hat{v} - \mu_{\hat{v}}\right)^2 \left(b - \mu_b\right)\right]}{\sigma_{\hat{v}}^2 \sigma_b} \right\} \\
&= \sigma_{\hat{v}}^2 \sigma_b \left\{ sk\left(\hat{v}\right) \rho\left(\hat{v}, b\right) - cosk\left(\hat{v}, b\right) \right\},
\end{aligned}$$

where BG1969 refers to Bohrnstedt and Goldberger (1969). Hence, a sufficient (and necessary) condition for $c_2 < 0$ is:

$$c_2 < 0 \iff \rho\left(\hat{v}, b\right) sk\left(\hat{v}\right) < cosk\left(\hat{v}, b\right) \tag{46}$$

and a sufficient (and necessary) condition for $c_2 > 0$ is:

$$c_2 > 0 \iff \rho\left(\hat{v}, b\right) sk\left(\hat{v}\right) > cosk\left(\hat{v}, b\right). \tag{47}$$

[23] This can be seen as follows:

$$\begin{aligned}
cov\left(Y, Y^2\right) &= E\left(Y^3\right) - \mu_{Y^2}\mu_Y = E\left(Y^3\right) - \mu_Y\left[E\left(Y^2\right) - \mu_Y^2 + \mu_Y^2\right] = E\left(Y^3\right) - \mu_Y\sigma_Y^2 - \mu_Y^3 \\
&= \frac{\left[E\left(Y^3\right) - \mu_Y^3 - 3\mu_Y\sigma_Y^2 + 3\mu_Y\sigma_Y^2\right]\sigma_Y^3}{\sigma_Y^3} - \mu_Y\sigma_Y^2 = sk(Y)\sigma_Y^3 + 3\mu_Y\sigma_Y^2 - \mu_Y\sigma_Y^2 = sk(Y)\sigma_Y^3 + 2\mu_Y\sigma_Y^2.
\end{aligned}$$

Sufficient Conditions for Bunching at the Top

To obtain sufficient conditions for Bunching at the Top, we need to find sufficient conditions for $c_1 > 0$, given that $c_2 < 0$. Using the results for c_1 from Equation (43), we obtain:

$$\begin{aligned} & c_1 > 0 \\ \Leftrightarrow & \frac{cov\left(v, \hat{v} - \frac{cov\left(\hat{v}, \hat{v}^2\right)}{var(\hat{v}^2)}\hat{v}^2\right)}{var\left(\hat{\epsilon}_1\right)} > 0 \\ \Leftrightarrow & \underbrace{cov\left(v, \hat{v}\right)}_{(1)} - \underbrace{cov\left(v, \hat{v}^2\right)}_{(2)} \underbrace{\frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)}}_{(3)} > 0. \end{aligned} \tag{48}$$

Term (1) of Inequality (48) is positive by assumption because we restrict ourselves to settings where partial bans become relevant (note that $A < 1$ from Proposition 5 is equivalent to $cov(v, \hat{v}) > 0$). A sufficient condition for Term (3) to be positive is that the skewness is positive, which follows from:

$$\begin{aligned} cov\left(\hat{v}, \hat{v}^2\right) > 0 & \Leftrightarrow 2\mu_{\hat{v}}\sigma_{\hat{v}}^2 + sk(\hat{v})\sigma_{\hat{v}}^3 > 0 \\ & \Leftrightarrow sk(\hat{v}) > -\frac{2\mu_{\hat{v}}}{\sigma_{\hat{v}}} \\ & \Leftarrow \boxed{sk(\hat{v}) > 0}. \end{aligned}$$

Next, we show that Inequality (48) holds under the assumptions made so far. When $cov\left(v, \hat{v}^2\right) \leq 0$, the inequality holds because Term (1) and Term (3) are positive. When $cov\left(v, \hat{v}^2\right) > 0$, we can derive an upper bound for Term (2) by rearranging Equation (44):

$$c_2 < 0 \iff cov\left(v, \hat{v}^2\right) < cov\left(v, \hat{v}\right) \frac{cov\left(\hat{v}^2, \hat{v}\right)}{var\left(\hat{v}\right)}. \tag{49}$$

Then, Inequality (48) holds if it holds after substituting the upper bound from Inequality (49) for Term (2):

$$\begin{aligned} cov\left(v, \hat{v}\right) - cov\left(v, \hat{v}\right) \frac{cov\left(\hat{v}^2, \hat{v}\right)}{var\left(\hat{v}\right)} \frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)} > 0 \\ \Leftrightarrow cov\left(v, \hat{v}\right)\left(1 - \rho^2_{\hat{v}^2, \hat{v}}\right) > 0, \end{aligned}$$

where $\rho^2_{\hat{v}^2, \hat{v}} \in (-1, 1)$ is the square of the correlation between $\hat{v}^2$ and $\hat{v}$. As we consider settings with $cov(v, \hat{v}) > 0$, this inequality holds.

To close the proof, note that $sk(\hat{v}) > 0$ in combination with Equation (46) implies that:

$$\boxed{\frac{cosk\left(\hat{v}, b\right)}{sk\left(\hat{v}\right)} > \rho\left(\hat{v}, b\right)}.$$

Hence, this condition and the condition that $\boxed{sk(\hat{v}) > 0}$ are sufficient for Bunching at the Top if $\hat{v} \in [0, \infty)$ and $cov(v, \hat{v}) > 0$.

Sufficient Conditions for Bunching at the Bottom

To obtain sufficient conditions for Bunching at the Bottom, we need to find sufficient conditions for $c_1 < 0$, given that $c_2 > 0$. Using the results for c_1 from Equation (48), we obtain:

$$c_1 < 0 \iff cov\left(v, \hat{v}\right) - cov\left(v, \hat{v}^2\right) \frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)} < 0. \tag{50}$$

Using Equation (45), we obtain that

$$cov\left(\hat{v}, \hat{v}^2\right) = \left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\sigma_{\hat{v}}^2.$$

Furthermore,

$$\begin{aligned}
cov\left(v, \hat{v}^2\right) &= E\left[\left(v - \mu_v\right)\left(\hat{v}^2 - \mu_{\hat{v}^2}\right)\right] \\
&= E\left(v\hat{v}^2\right) - \mu_{\hat{v}^2}\mu_v - \mu_v E\left(\hat{v}^2\right) + \mu_v\mu_{\hat{v}^2} \\
&= \frac{\left[E\left(v\hat{v}^2\right) - \mu_v E\left(\hat{v}^2\right) - 2E\left(v\hat{v}\right)\mu_{\hat{v}} + 2\mu_{\hat{v}}^2\mu_v + 2E\left(v\hat{v}\right)\mu_{\hat{v}} - 2\mu_{\hat{v}}^2\mu_v\right]\sigma_{\hat{v}}^2\sigma_v}{\sigma_{\hat{v}}^2\sigma_v} \\
&\overset{(*)}{=} cosk(\hat{v}, v)\sigma_{\hat{v}}^2\sigma_v + 2E\left(v\hat{v}\right)\mu_{\hat{v}} - 2\mu_{\hat{v}}^2\mu_v \\
&= cosk(\hat{v}, v)\sigma_{\hat{v}}^2\sigma_v + 2\mu_{\hat{v}}\left[E\left(v\hat{v}\right) - \mu_{\hat{v}}\mu_v + \mu_{\hat{v}}\mu_v - \mu_{\hat{v}}\mu_v\right] \\
&= cosk(\hat{v}, v)\sigma_{\hat{v}}^2\sigma_v + 2\mu_{\hat{v}}cov\left(v, \hat{v}\right) \\
&= \left[cosk(\hat{v}, v)\sigma_{\hat{v}} + 2\mu_{\hat{v}}\rho\left(v, \hat{v}\right)\right]\sigma_{\hat{v}}\sigma_v,
\end{aligned}$$

where in $(*)$ we use that $cosk(\hat{v}, v) = E\left(v\hat{v}^2\right) - 2E\left(v\hat{v}\right)\mu_{\hat{v}} - 2E\left(\hat{v}^2\right)\mu_v + 2\mu_{\hat{v}}^2\mu_v$. Hence, we can rewrite Inequality (50) as follows

$$\begin{aligned}
& cov\left(v, \hat{v}\right) - cov\left(v, \hat{v}^2\right)\frac{cov\left(\hat{v}, \hat{v}^2\right)}{var\left(\hat{v}^2\right)} < 0 \\
\Leftrightarrow\ & cov\left(v, \hat{v}\right) - \left[cosk(\hat{v}, v)\sigma_{\hat{v}} + 2\mu_{\hat{v}}\rho\left(v, \hat{v}\right)\right]\sigma_{\hat{v}}\sigma_v\frac{\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\sigma_{\hat{v}}^2}{var\left(\hat{v}^2\right)} < 0
\end{aligned}$$

$$\Leftrightarrow \left[cosk(\hat{v},v)\sigma_{\hat{v}} + 2\mu_{\hat{v}}\rho\left(v,\hat{v}\right)\right]\sigma_{\hat{v}}\sigma_{v}\frac{\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\sigma_{\hat{v}}^{2}}{var\left(\hat{v}^{2}\right)} > cov\left(v,\hat{v}\right)$$

$$\Leftrightarrow \left[cosk(\hat{v},v)\sigma_{\hat{v}} + 2\mu_{\hat{v}}\rho\left(v,\hat{v}\right)\right]\frac{\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\sigma_{\hat{v}}^{2}}{var\left(\hat{v}^{2}\right)} \overset{(*)}{>} \rho\left(v,\hat{v}\right)$$

$$\Leftrightarrow \left[\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)}\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\frac{\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\sigma_{\hat{v}}^{2}}{var\left(\hat{v}^{2}\right)} > 1$$

$$\Leftrightarrow \left[\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)}\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right] > \frac{var\left(\hat{v}^{2}\right)}{\sigma_{\hat{v}}^{2}}$$

$$\Leftrightarrow \left[\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)}\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right]\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}}\right] - \frac{\sigma_{\hat{v}^2}^{2}}{\sigma_{\hat{v}}^{2}} > 0$$

$$\Leftrightarrow \underbrace{\left[\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)}\sigma_{\hat{v}} + 2\mu_{\hat{v}} + \frac{\sigma_{\hat{v}^2}}{\sigma_{\hat{v}}}\right]}_{(1)}\underbrace{\left[sk(\hat{v})\sigma_{\hat{v}} + 2\mu_{\hat{v}} - \frac{\sigma_{\hat{v}^2}}{\sigma_{\hat{v}}}\right]}_{(2)} + \underbrace{\sigma_{\hat{v}^2}\left[\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)} - sk(\hat{v})\right]}_{(3)} > 0,$$

where in $(*)$ we use that $cov\left(v,\hat{v}\right) > 0$. The last inequality holds if $(1) < 0$, $(2) < 0$, and $(3) > 0$. Note that $(3) > 0$ follows from $c_2 > 0$. Furthermore, $(1) < 0$ if:

$$\frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)} < -2\frac{\mu_{\hat{v}}}{\sigma_{\hat{v}}} - \frac{\sigma_{\hat{v}^2}}{\sigma_{\hat{v}}^{2}} =: k, \tag{51}$$

where $k < 0$. From $(3) > 0$ and $(1) < 0$ it also follows that $(2) < 0$, which can be seen as follows:

$$sk(\hat{v}) \overset{(3)>0}{<} \frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)} \overset{(1)<0}{<} -2\frac{\mu_{\hat{v}}}{\sigma_{\hat{v}}} - \frac{\sigma_{\hat{v}^2}}{\sigma_{\hat{v}}^{2}} < -2\frac{\mu_{\hat{v}}}{\sigma_{\hat{v}}} + \frac{\sigma_{\hat{v}^2}}{\sigma_{\hat{v}}^{2}}. \tag{52}$$

To close the proof, note that $sk(\hat{v}) < 0$ in combination with Equation (47) implies that:

$$\boxed{\frac{cosk\left(\hat{v},b\right)}{sk\left(\hat{v}\right)} > \rho\left(\hat{v},b\right)}.$$

Hence, this condition and the condition that $\boxed{sk(\hat{v}) < \frac{cosk(\hat{v},v)}{\rho\left(v,\hat{v}\right)} < k < 0}$ are sufficient for Bunching at the Bottom if $\hat{v} \in [0,\infty)$ and $cov(v,\hat{v}) > 0$.

B Optimal Non-Linear Taxation in the Light Bulb Market

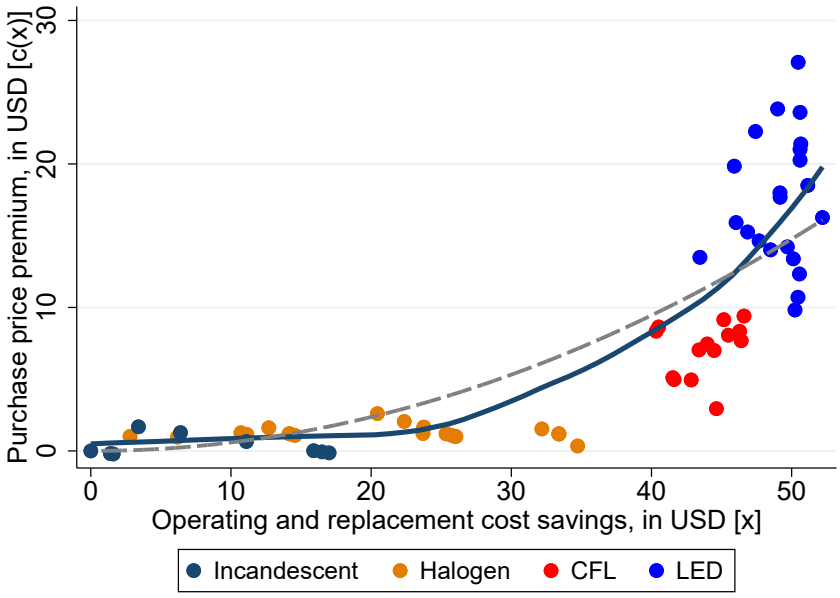
We first explain how we approximate the cost function for energy efficiency and consumers' normative valuations v. In a subsequent step, we present the aggregate welfare effects of the optimal linear and non-linear subsidy on energy efficiency, relative to no taxation.

Our supply data stems from a price comparison service, *geizhals.de*, which reports the cheapest price of a product offered on an online website. We focus on light bulbs that are typically purchased by households. In particular, we consider bulbs with an energy intensity of 25 to 75 Watt-equivalents and a warm light color of around 2700 Kelvin. To reduce the impact of branding effects, we focus on bulbs produced by one of the two large manufacturers, *Osram* and *Philips*, that offer bulbs both in the European Union and the United States. As in Allcott and Taubinsky (2015a), we express all prices in 2012 US dollars (USD) and collect product prices during that year. Some LED and CFL bulbs enter the market after 2012: in these cases, we extrapolate their 2012 price based on their aggregate annual price trends, which imply a 20% and 10% price decrease per annum for LED and CFL bulbs, respectively. For every bulb, we determine the operating and replacement cost (ORC) to consume 8.000 hours of light over eight years, which corresponds to three hours per day, assuming electricity prices of 0.1 USD per kWh (Allcott and Taubinsky, 2015a).

Based on this data, we determine the purchase price premiums and ORC savings relative to the most electricity-intensive bulb. In the following, we use ORC savings as the measure of attribute level q, i.e., of energy efficiency. Figure 2 plots the price premiums against ORC savings, which corresponds to the cost function $c(q)$ in our model. The least energy inefficient, yet cheapest, bulbs are incandescent bulbs, followed by halogen, CFL and LED bulbs. The cost curve is convex, which reflects that a one unit increase in ORC savings becomes increasingly more expensive as the level of energy efficiency increases. In 2012, the most energy efficient LED bulbs sold at a price premium of around 30 USD and yielded cost savings of about 50 USD over the course of eight years, compared to the most energy inefficient incandescent bulbs.

We use the elicitation of time preferences by Allcott and Taubinsky (2015a) to determine individual-specific discount factors. We assume that all other factors that influence normative valuations do not vary by participant and thus merely constitute a scaling factor. This assumption allows us to calibrate valuations to match the supply function from Figure 2. In particular, we set valuations to $v = s \cdot D(\delta)$, where s is a scaling factor that ensures that consumers demand every product variety offered on the market. To illustrate, consider a consumer with a discount rate of $\delta = 20\%$ and annual operating cost of $1/8$ USD for eight years, which results in ORC savings of 1

Figure 2: Energy Efficiency Cost Function in the Light Bulb Market



Note: Price premiums, as well as operating and replacement cost savings are determined relative to the most electricity intensive bulb. Operating and replacement cost assume eight years of total usage (8.000 hours) and an electricity price of 0.1 USD per kWh, as in Allcott and Taubinsky (2015a). The solid line plots predictions from a local linear regression (bandwidth: 9), while the dashed line plots predictions from a regression on quadratic terms.

USD. For that consumer, the normative valuation of 1 USD in ORC savings is then $D(\delta) = (1 + 1/(1+\delta) + \cdots + 1/(1+\delta)^7) \cdot (1/8) = 0.58$ USD. We consider this approach as a useful approximation of individuals' normative preferences v that isolates one source of heterogeneity in v and is consistent with observable market behavior.[24]

We assume a quadratic cost function and estimate it based on the data from Figure 2. We derive consumers' choices in five scenarios: the absence of a corrective tax, the presence of the optimal linear tax, the optimal non-linear tax a) under full information, b) under a first-order approximation of the conditional bias, and c) under a second-order approximation of the conditional bias. The aggregate welfare effects are presented in Table 1.

[24]We set the scaling factor to 0.6, which ensures that the largest perceived valuations are equal to the highest gradient of the cost function from Figure 2. In addition, we impose some consistency restrictions on the data. Starting with a sample of 633 individuals with non-censored valuations in the treatment groups, we drop all observations with missing values on biases and discount rates (23 observations). We also drop observations where the elicitation of time preferences does not yield a discount factor between 0 and 1 (50 observations) and where biases or perceived valuations are above below the 1 or above the 99 percentile (14 observations). In addition, we drop all observations where perceived valuations would be negative (57 observations), which leaves us with 489 observations for our numerical example.

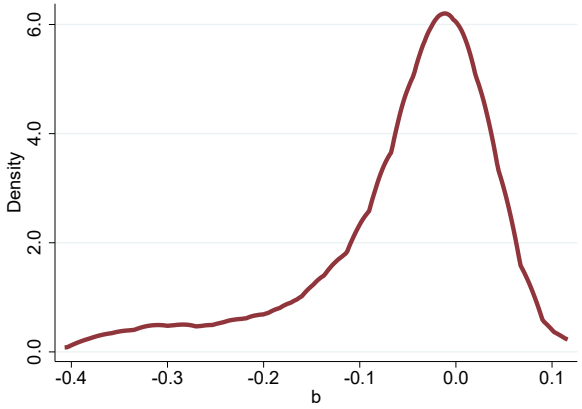
Table 1: Welfare Implications of Taxation

	Mean welfare, in USD/bulb	Mean welfare gain over status quo, in EUR/bulb	Mean welfare gain relative to linear tax, in %
Status quo (no tax)	3.79	0.00	–
Linear tax	3.91	0.12	0
Non-linear tax (first-order approx.)	3.96	0.16	36
Non-linear tax (second-order approx.)	3.97	0.17	44
Non-linear tax (full information)	3.98	0.19	57
Welfare effects under a first-order approximation, (wrongly) setting $\rho = 0$			
Non-linear tax (first-order approx, $\rho = 0$)	3.95	0.16	31

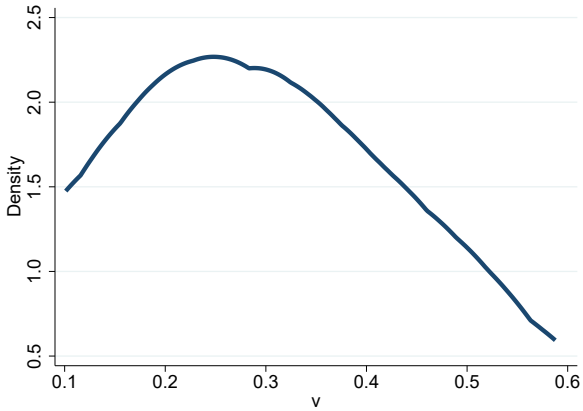
Note: Mean welfare is calculated under the optimal linear and non-linear tax schedules using the joint distribution of perceived valuations and biases, as well as the cost function estimated in Section 5 and Appendix Section B. "Non-linear tax (first-order approx, $\rho = 0$)" implements the optimal non-linear tax based on the first-order approximation to the expected bias, (wrongly) setting $\rho = 0$.

Figure 3: Distribution of bias and normative valuations

(a) Distribution of bias b



(b) Distribution of normative valuation v



Notes for Figure a) and b): Densities estimated via kernel density estimation (Epanechikov kernel, bandwidth: 0.03 and 0.1, respectively).