Attending to Inattention: Identification of Deadweight Loss under Non-Salient Taxes

Giacomo Brusco* and Benjamin Glass† ‡

August 8, 2019

Abstract

Recent developments in behavioral public economics have shown that heterogeneous biases prevent point identification of deadweight loss. We replicate this result for an arbitrary (closed) consumption set, whereas previous results on heterogeneous attention focused on binary choice. We find that one can bound the efficiency costs of taxation based on aggregate features of demand. When individuals have linear demand functions, the bounds for deadweight loss are easy to calculate from linear regressions.

*Economics, University of Michigan at Ann Arbor

†Economics, Pennsylvania State University at University Park

‡The work in the following pages would have been impossible without the thought-provoking conversations we had with James Hines, Joel Slemrod, Dmitry Taubinsky, Alex Rees-Jones, Jeremy Fox, Tilman Börgers, Ying Fan, Ugo Troiano, William Boning, Luis Alejos, Tejaswi Velayudhan, Aristos Hudson, Benjamin Lockwood, Sarah Moshary, Alessandro Lizzeri, Amine Ouazad, the participants of the 2017 NTA conference, the participants of the 2018 Quebec Political Economy Conference, and the participants of the Public Finance, Theory, and IO seminars at the economics department of the University of Michigan. We warmly thank all of them for their invaluable advice. Last but not least, a special thanks goes to Johannes Spinnewijn in quality of associate editor at the Journal of Public Economics and to anonymous reviewers at the Journal of Public Economic Theory for their insightful comments that reshaped and refocused the scope of this paper. Any mistakes are, of course, our own.

1 Introduction

Taxing a good results in a loss of economic efficiency whenever it distorts equilibrium behavior away from the Pareto optimum. To the extent that agents do not notice a tax, the burden of the tax is exacerbated by the fact that agents cannot adjust the behavior to protect themselves from the tax. However, the burden of the tax in excess of government revenue, or deadweight loss, is mitigated when agents do not pay attention to the tax: if consumers pay the tax without noticing it, they are effectively transferring some of their income to the government in a lump sum. Chetty, Looney, and Kroft (2009, henceforth CLK) were the first to make these points. In addition to their theoretical contributions, they also showed that consumers in the U.S., where sales tax is applied at the register rather than included on the prices displayed on shelves (or, "sticker prices"), tend to under-react to sales taxes.

While CLK (2009) focuses on the case of homogeneous attention, recent work by Taubinsky and Rees-Jones (2018, henceforth TRJ) has noted that introducing the possibility of heterogeneous attention may prevent the computation of deadweight loss from aggregate data. If each person faced a different tax rate when buying a certain good, understanding welfare effects would require us to study not only aggregate demand, but the demand of every individual. Imposing a high tax on low elasticity individuals and a low tax on high elasticity individuals would have a very different effect on welfare than doing the opposite. A similar reasoning applies when all agents face the same tax rate, but perceive different tax rates. TRJ (2018) find that allowing for heterogeneous attention introduces an issue of allocative inefficiency that is normally absent from the study of the welfare effects of taxation. In a world of heterogeneous attention, there is no guarantee that the individuals who end up consuming the good are the ones who value it the most.

TRJ (2018) make these points in a binary choice model. This is well-suited to their experiment, in which people are choosing whether or not to buy a certain object, but their proof does not generalize trivially. Given the predominance of continuous choice settings in much of the literature on tax salience, including CLK's seminal paper, this motivates us to study the issue further.

We begin by developing a model of choice under misperceived prices with an arbitrary closed consumption set, and develop our welfare measure: compensating variation due to the tax, net of tax revenue. Bernheim and Rangel (2009) laid the foundations of welfare analysis with behavioral agents. Our model is similar to the models of CLK (2007, 2009) and TRJ (2018), but we slightly modify the treatment of income effects, along the lines of Gabaix's (2014) model of rational inattention. In the absence of income effects, our model of choice for an individual agent is essentially equivalent to the model in CLK (2009), except that we allow for arbitrary consumption sets. Our model is also similar to the model of Chetty (2009), but for the fact that we specify a particular way in which behavioral agents maximize utility. While this does not impose severe restrictions on behavior, it offers a useful framework when we move on to identification. We confirm that some of the major results in CLK (2009) and TRJ (2018) hold quite broadly: inattention to taxes increases the size of the loss in consumer surplus, but decreases the size of deadweight loss; attention heterogeneity amplifies deadweight loss, and invalidates CLK's sufficient statistic approach.

The main contribution of this paper is to generalize TRJ's non-identification result to an arbitrary closed choice set. We show that an econometrician who only observes aggregate consumption data can only determine the true value of aggregate deadweight loss to lie on an interval. These bounds were first noted by TRJ (2018) in their proposition A.2. We find these bounds hold generally and propose to use them as a novel empirical tool.

The lower bound for deadweight loss is the calculation one would perform in the case of a representative consumer. Since the loss in efficiency is a convex function of the perceived tax rate, the calculation of deadweight loss from *one* perceived tax-inclusive price consistent with aggregate demand will generically underestimate deadweight loss. Heterogeneity in tax salience creates heterogeneity in perceived net-of-tax prices, which creates an allocative inefficiency across consumers. As the calculation with a representative consumer only accounts for inefficiency from *aggregate* foregone consumption due to the tax, it will underestimate excess burden. However, in the case in which all agents pay the same amount of attention to the tax, there is no allocative inefficiency between consumers, and so performing the calculation as with a representative consumer yields the correct value for deadweight loss. The formula for this lower bound to deadweight loss is an extension of formulas provided by CLK (2009) and TRJ (2018).

Following TRJ (2018), we obtain an upper bound for deadweight loss by letting the econometrician assume that tax salience has support on a known bounded non-negative interval. The upper bound comes from maximizing perceived price heterogeneity, again exploiting the convexity of deadweight loss with respect to the perceived tax. This is achieved by positing that agents have either zero or maximal salience. Generalizing introduces two additional considerations in calculating the upper bound for deadweight loss. One, a distribution yielding the upper bound for deadweight loss assigns high tax salience precisely where it will "hurt" most: to those agents whose particular preferences yield maximal deadweight loss from that agent relative to the change in consumption of that agent. This distribution allocates high tax salience to those agents who have more convex demand curves, keeping the aggregate change in quantity demanded constant. Two, deadweight loss is maximized for a given aggregate demand if any agent with multiple optimal decisions at the perceived price consumes the highest amount consistent with their preference when they perceive low prices, whereas they consume the lowest amount consistent with their preferences when they perceive high prices.[1] This is because heterogeneity in perceived prices permits different equilibria with the same sticker price, tax rate, and aggregate consumption, yet yielding different values of deadweight loss due to different distributions of consumption among individuals.

Our approach to compute the upper bound forces the econometrician to impose limits on the possible values of attention – something that the empirical literature has not settled yet and might be highly context-dependent. While it might seem natural to assume that attention varies between zero and one, many papers have found evidence of salience above one – see for instance Allcott and Taubinsky (2015). In theory, one might allow for unlimited (positive) tax salience, under regularity conditions that avoid the possibility of unlimited distortion to consumer behavior.[2]

Our general results follow the work of TRJ (2018), and rely on the fact that the distribution of preferences is independent of taxes and prices, but that the distribution of salience might change depending on the value of the tax. Indeed, we show that, even assuming that the distribution of preferences is entirely known to the econometrician and invariant to observables, one cannot identify deadweight loss.

We then turn to the special case in which demand is linear in a relevant range where consumption is positive. This case is of particular interest for three reasons: first, its ease of applicability; second, its special relationship to the second order approximation of deadweight loss; and third, its illustrative value in the kind of problems we can face in identifying deadweight loss. In the case of binary choice, non-identification ultimately comes from the possibility that taxes and attention to taxes are not independent of each other. Although the experimental evidence in TRJ (2018) suggests that attention varies with how large the tax is, if one assumed away this possibility we could identify a full distribution of responsiveness to both taxes and sticker prices using existing models of discrete choice with random coefficients, as in Masten (2017) and Fox (2017). This estimated distribution could then be used to yield a point-estimate of deadweight loss. In the case of linear demand, instead, even assuming that attention and taxes are independent would not help identify deadweight loss with aggregate data beyond the bounds described above.

Our results complement a growing literature on tax salience. Rosen (1976) does not find evidence of limited tax salience, but besides CLK (2009) and TRJ (2018), Finkelstein (2009), Gallagher and Muehlegger (2011), and Goldin and Homonoff (2013) all find strong evidence of dramatically limited tax salience. Most of this literature looks at sales taxes in the U.S., as they lend themselves very credibly to a story about lack of salience. However, work on salience has also looked at other settings: Finkelstein (2009) studies car tolls; Weber and Schram (2016) study whether income taxes being remitted by the employer or the employee affects differently people's attitude towards public spending and the burden of the tax;[3] Morone, Nemore and Nuzzo (2018) explore a similar question in the context of a double-auction market. Blake, Moshary and Tadelis (2017) study how people react differently to back-end and upfront fees in online purchases; and Bradley and Feldman (2018) study how changes in the disclosure of ticket taxes affect the demand for airlines. As we mentioned above, most of these empirical papers, as well as other theoretically focused papers like Goldin (2015), use models where the choice set is continuous or mixed discrete-continuous.

The paper proceeds as follows. In section 2, we develop a general model of choice under misperceived prices. Once we have replicated some of the major theoretical results in CLK (2009) and TRJ (2018), we

[1] TRJ (2018) do not deal with cases like this because they restrict attention to non-atomic distributions of willingness to pay, so almost every agent has a unique choice that is perceived to maximize utility.

[2] One might assume that consumer surplus is uniformly bounded to ensure that the upper bound for deadweight loss is finite even if the upper bound for tax salience is infinite.

[3] Interestingly, assuming some agents face credit constraints as in Boadway, Garon, and Perrault (2018), also breaks traditional optimal tax theory.

shift focus to identification of deadweight loss. Section 3 lays out the main results of our paper, establishing the non-identification result and the bounds. Section 4 focuses on the special case in which demand is linear, and provides a straightforward way to compute our bounds in the context of linear models. Section 5 concludes. Proofs and other minor results are relegated to the online appendix.

2 Choice and Deadweight Loss under Non-Salient Taxes

This section describes the theoretical model and results that underlie the rest of this paper. Many of our results here simply mirror previous literature, but we make slightly different modeling choices. The main modeling challenge in dealing with misperceived prices is to allow for the misperception of prices while keeping agents financially solvent. CLK (2007, 2009) assume that one good "absorbs" all optimization mistakes. In contrast our approach, inspired by parts of the model in Gabaix (2014), has agents conjecture a certain income such that they end up consuming on their true budget constraint when presented with the relative prices they perceive. While this framework preserves all results of interest from CLK (2009) and TRJ (2018), we find that our approach eases exposition while freeing the researcher from having to make ad-hoc assumption about which good (or goods) absorb optimization mistakes. It should be noted that while we do generalize the model to include multiple taxed and non-taxed goods in the online appendix, in the body of the paper agents will face a choice over two goods, only one of which is subject to tax.

The agent has a closed consumption set $X = X^T \times \mathbb{R}_+ \subseteq \mathbb{R}^2_+$. She also has a choice function for the taxed good, $q(\bar{p}, p^{NT}, \tau, W)$, with $(\bar{p}, p^{NT}) \in \mathbb{R}^2_{++}$, where $\bar{p}$ and p^{NT} are respectively the sticker price of the taxed and non-taxed good, $\tau \in \mathbb{R}$ is the sales tax on the taxed good, and W is the income of the agent.[4] We express taxes as if they were specific, so that $\bar{p} + \tau$ is the tax-inclusive price of the taxed good.

The agent has a continuous and strictly monotonic utility function $u(q, q^{NT})$, where q denotes generic consumption of the taxed good. The choice vector function $\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W) = (q(\bar{p}, p^{NT}, \tau, W), q^{NT}(\bar{p}, p^{NT}, \tau, W)) \in X$ meets two requirements. One, the agent spends all available income:[5]

$$(\bar{p} + \tau)q(\bar{p}, p^{NT}, \tau, W) + p^{NT} q^{NT}(\bar{p}, p^{NT}, \tau, W) = W. \tag{1}$$

Two, the agent correctly optimizes in the choice of all consumption bundles when there is no tax:

$$\boldsymbol{q}(\bar{p}, p^{NT}, 0, W) \in \underset{\{\bar{p}*q + p^{NT}*q^{NT} \leq W\}}{\arg\max} u(q, q^{NT}). \tag{2}$$

It turns out that this model is quite general, in the sense that it rules out very few possible behaviors. Indeed, this model is equivalent to one in which agents pick rationally given a perceived price p^s, and conjecture themselves an income W^s so that they satisfy their true budget constraint at their perceived price. Proposition 2 and subsequent work in the online appendix shows that, under weak convexity assumptions on preferences, one can find a pair (p^s, W^s) to satisfy equations 1 and 2 for any choice function $\boldsymbol{q}(\cdot)$.

We now want some measure of the incidence of the tax on the consumer. For concreteness, we consider the compensating variation due to the tax with complete pass-through, defined as:

$$\Delta CS \equiv \inf\{\Delta W |\ u(\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W + \Delta W)) \geq u(\boldsymbol{q}(\bar{p}, p^{NT}, 0, W))\}.$$

[4] We implicitly restrict consideration to sticker prices, taxes, and income such that q$(\bar{\boldsymbol{p}}, \tau, W)$ is well-defined at those values.

[5] When we consider multiple non-taxed goods in the online appendix, we also have agents optimally choose q^{NT} given their choice of q.

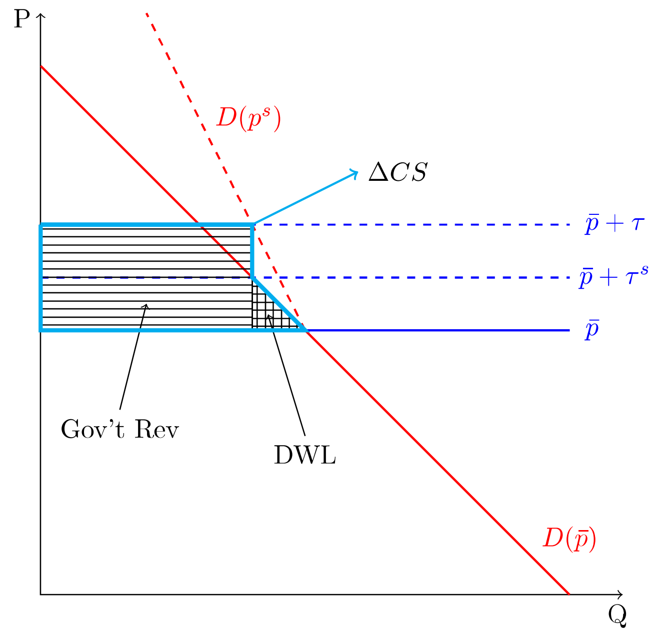


Figure 1: Welfare effects from the imposition of a non-salient tax.

In words, the change in consumer surplus is the greatest lower bound of the amount of money the agent requires to achieve the utility reached before the imposition of the tax. Online appendix proposition 3 shows that the change in consumer surplus can be written as the sum of what consumer would have to be compensated if the tax-inclusive price were *actually* p^s and the income they "lost" due to their inattention:

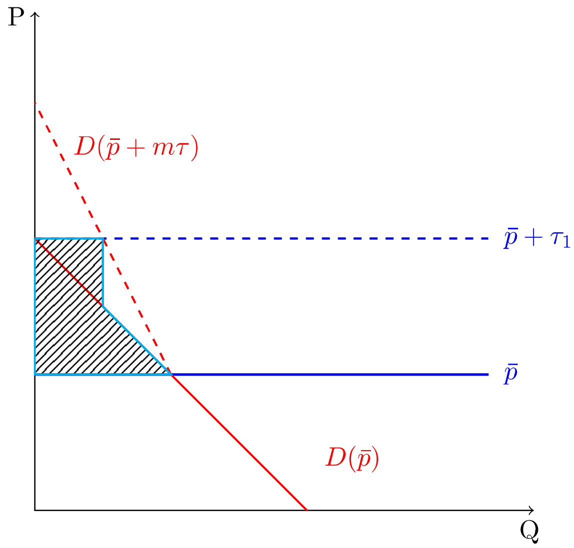
$$\Delta CS = \underbrace{(\bar{p}+\tau-p^s)h(p^s)}_{\text{Income lost}} + \underbrace{e(p^s)-e(\bar{p})}_{\Delta CS \text{ under } p^s}. \tag{3}$$

This representation, which is graphically illustrated in figure 1, readily gives us two interesting results. First, as noted by CLK (2007, 2009), if $p^s \in [\bar{p}, \bar{p}+\tau]$, then the consumer will be worse off than if she paid attention to the tax:

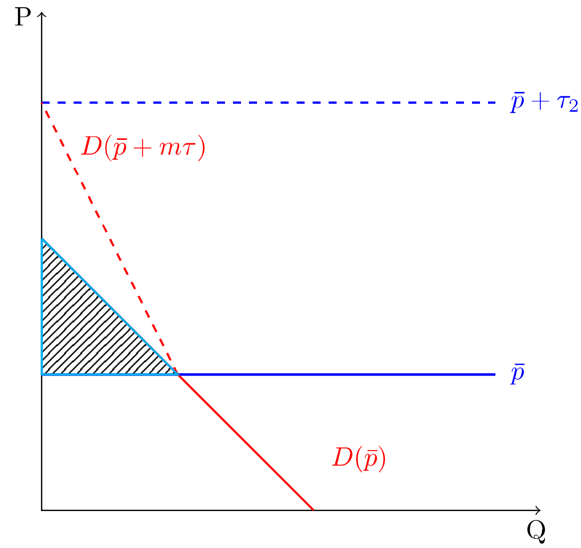
$$\begin{aligned}\Delta CS &= (\bar{p}+\tau-p^s)h(p^s) + \int_{\bar{p}}^{p^s} h(p)dp \\ &\geq \int_{p^s}^{\bar{p}+\tau} h(p)dp + \int_{\bar{p}}^{p^s} h(p)dp \\ &= \int_{\bar{p}}^{\bar{p}+\tau} h(p)dp.\end{aligned}$$

Second, a consumer can be made better off by an increase in the tax, which we do not believe previous literature to have noted. This is because a tax increase can induce inattentive agents to reduce their consumption of the taxed good to zero, improving welfare for consumers who would have already avoided consuming the taxed good had they been attentive to the tax.[6] We provide a graphical example in figure 2.

[6]More formally, consumption does not necessarily have to be reduced to zero for the consumer to be made better off. The loss of consumer surplus decreases in τ whenever the tax is sufficiently high such that consumption is sufficiently small (but positive).



(a) Under τ_1 the consumer loses more than total CS



(b) Under τ_2 the consumer loses exactly total CS

Figure 2: The consumer in (a) is worse off than in (b), although she is subject to a lower tax

To obtain deadweight loss, we need to adjust ΔCS for the change in tax revenue:[7]

$$dwl \equiv \Delta CS - \tau q(\bar{p}, p^{NT}, \tau, W + \Delta CS) = e(p^s) - e(\bar{p}) - (p^s - \bar{p})q(\bar{p}, p^{NT}, W + \Delta CS). \tag{4}$$

Note that deadweight loss is exactly as if the agent was correct that the tax-inclusive price were p^s.

To introduce heterogeneity, let $i \in \mathcal{I}$ index consumers. Each consumer is characterized by her perception of the price of the taxed good, p_i^s, type, θ_i, standing in for her preferences $\succeq_{\theta_i}$ and income W_{θ_i}, and tie-breaking parameter ζ_i, which we need for technical reasons.[8] These consumer-specific parameters are distributed according to $F^*_{p^s,\theta,\zeta}$. Each agent has a choice function for the taxed good, satisfying:

$$q(\bar{p}, p^{NT}, \tau, W_{\theta_i}; \theta_i, \zeta_i) = q(p_i^s, W_i^s; \theta_i, \zeta_i) \in \{q | \exists q^{NT} : (q, q^{NT}) \succeq_{\theta_i} \boldsymbol{q'} \ \forall \boldsymbol{q'} \in X : (p_i^s, p^{NT}) * \boldsymbol{q'} \leq W_i^s\},$$

where p_i^s and W_i^s are determined as in the model of section 2, with corresponding expenditure function $e(p_i^s; \theta_i)$. If demand is single-valued, letting us ignore the tie-breaking parameter ζ, total deadweight loss is:

$$DWL = \int_{p_i^s, \theta_i} [e(p_i^s; \theta_i) - e(\bar{p}; \theta_i)] - (p_i^s - \bar{p})q(p_i^s; \theta_i) dF^*_{p^s,\theta}(p_i^s, \theta_i).$$

This paper emphasizes continuous choice, as the binary choice case is worked out in TRJ (2018). We momentarily assume that h is continuously differentiable with respect to its own price, and that p^s is continuously differentiable with respect to τ. Then tax salience at a tax rate of zero is $m = \frac{\partial p^s}{\partial \tau}|_{(\bar{p},0)}$.[9] In line with existing deadweight loss analyses, we can consider a second-order approximation, letting us characterize

[7] We maintain the convention that deadweight loss is a generically positive value.

[8] ζ_i acts as a tie-breaker among bundles that could all have been chosen: choices do not necessarily reflect true preferences when agents misperceive prices, and agents might appear indifferent between choices that do not actually yield the same ex-post utility. This is in sharp contrast with the neo-classical model, where the actual choice that one selects among indifferent bundles has no impact on consumer surplus. But in this model of choice, different values of conjectured income might yield different choices, even given the same preferences, perceived prices, and true income.

[9] Formally, the claim is that there exist (p^s, W^s) such that $\frac{\partial p^s}{\partial \tau}$, where the derivative is taken while the consumer is being compensated, is well defined. If $\frac{\partial h}{\partial p}(\bar{p}) \neq 0$, then the Inverse Function Theorem implies that $\frac{\partial p^s}{\partial \tau} = \frac{\frac{\partial q}{\partial \tau} + \frac{\partial q}{\partial W}\frac{\partial \Delta CS}{\partial \tau}}{\frac{\partial h}{\partial p}}$. If $\frac{\partial h}{\partial p} = 0$ in a neighborhood around $\bar{p}$, then set $\frac{\partial p^s}{\partial \tau}|_{\tau=0} = 0$ and $\frac{\partial \Delta CS}{\partial \tau} = -\frac{\frac{\partial q}{\partial \tau}}{\frac{\partial q}{\partial W}}$.

our object of interest in terms of first derivatives:

$$DWL \approx -\frac{\tau^2}{2} \int_{p_i^s, \theta_i} m_i^2 \frac{\partial h(\bar{p}; \theta_i)}{\partial p} dF^*_{p^s, \theta}(p_i^s, \theta_i). \tag{5}$$

This process of aggregation makes apparent two important points that confirm the analysis of TRJ (2018) extends naturally to continuous choice. First, allowing for attention heterogeneity introduces an issue of allocative inefficiency, as it is no longer guaranteed that consumers who value some units of the taxed good the most will be the ones who end up purchasing those units. It should be noted that we are assuming throughout this paper that supply is perfectly elastic, and so tax increases will be reflected one-for-one in the after-tax price. This is not really an issue of concern: as TRJ (2018) note, all that is needed to generalize this to an arbitrary supply function is to account for the change in sticker price and the change in profits to suppliers. Nonetheless, as in TRJ's work, it is interesting to deviate for a moment from this assumption, to consider what happens to aggregate deadweight loss when supply is perfectly inelastic. In that case, we can use appendix proposition 4 to show that

$$DWL \approx -\frac{\tau^2}{2} \int_{m_i, \theta_i} \left[m_i^2 \frac{\partial q_i}{\partial p} dF_{m,\theta}(m_i, \theta_i) - \frac{(\int_{m_i, \theta_i} m_i \frac{\partial q_i}{\partial p})^2}{\int_{m_i, \theta_i} \frac{\partial q_i}{\partial p} dF_{m,\theta}(m_i, \theta_i)} \right] \geq 0,$$

and $DWL = 0$ when attention is homogeneous, i.e. $m_i = m \; \forall i$. Thus, a non-salient tax may yield excess burden even without changing the equilibrium quantity, due to its effects on allocative efficiency.

Second, allowing for heterogeneous attention introduces a serious problem of identification, as neither aggregate price responsiveness $\int_{p_i^s, \theta_i} \frac{\partial h(\bar{p}; \theta_i)}{\partial p} dF^*_{p^s, \theta}(p_i^s, \theta_i)$ nor aggregate tax responsiveness $\int_{p_i^s, \theta_i} m_i \frac{\partial h(\bar{p}; \theta_i)}{\partial p} dF^*_{p^s, \theta}(p_i^s, \theta_i)$ are sufficient statistics for the deadweight loss in equation 5. These points effectively extend some of TRJ's (2018) major results to continuous choice. In the next section, we formalize this non-identification result with an arbitrary choice function, and show how one might bound deadweight loss with mere information on aggregate parameters.

3 Non-Identification with Aggregate Data

This section discusses to what degree one can infer deadweight loss from aggregate choice data. Our results generalize TRJ's work on binary choice to an arbitrary choice set. We find this to be of interest for two reasons. First, many papers dealing with tax salience, including the seminal paper by CLK (2009), operate in a continuous choice setting. Second, while point identification is impossible using aggregate parameters, we can still provide tight bounds based solely on aggregate (or average) quantities.

For simplicity, we assume that the econometrician already knows the distribution of preference types, but this should not be considered a limiting assumption. Consumer preferences can be identified with sticker price variation when there are no (non-salient) taxes. Regardless, even when the econometrician can fully observe the true distribution of preferences, and how much aggregate consumption there is at every tax level, she still cannot infer the exact value of deadweight loss. However, we provide a lower bound and an upper bound for deadweight loss. The lower bound is achieved by assuming that all agents perceive the same tax-inclusive price, i.e. assume there is no attention heterogeneity. The upper bound for deadweight loss is achieved by imposing maximal attention heterogeneity. Since the data do not reveal the individual variation in tax salience, one cannot point identify deadweight loss from aggregate data. Deadweight loss can take on any value between the upper and lower bounds.[10]

The results in this section are described as if all agents face the same sales tax. Also, since we are considering the problem of identification with aggregate demand, we assume there are no income effects. This is because even the standard model with fully salient taxes requires strong restrictions on income effects in order to achieve identification with aggregate data, as the same income can yield different consumption bundles whenever there are several conjectured incomes that solve the agent's problem. Suppressing income

[10]This claim follows by taking any weighted average of the distributions of parameters yielding upper and lower bounds of deadweight loss.

and price for the non-taxed good, we denote the consumption function for agent i with type θ_i and perceived tax-inclusive price p_i^s for the taxed good by $q(p_i^s;\theta_i,\zeta_i)$. However, all of these results follow if one reinterprets $q(p_i^s;\theta_i,\zeta_i)$ as the *compensated* choice of agent i.

To ensure integrability, we assume the econometrician knows that $F^*_{p^s}$ has support with lower bound greater than zero, and so only considers marginal distributions of subjective prices bounded above zero. The econometrician observes aggregate demand:

$$\int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF^*_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).$$

Deadweight loss for an individual i is a function of their expenditure function $e(p)$ and prices via:

$$dwl(p_i^s;\theta_i,\zeta_i) = e(p_i^s;\theta_i) - e(\bar{p};\theta_i) - (p_i^s - \bar{p})q(p_i^s;\theta_i,\zeta_i).$$

We are interested in aggregate deadweight loss:

$$DWL = \int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF^*_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).$$

The problem of identification is to find conditions for which any joint distribution $F_{p^s,\theta,\zeta}$ of (p^s,θ,ζ), as a function of observable variables $\bar{p}$ & τ, satisfying these conditions and such that aggregate demand is rationalized; that is, such that for any observed values of observable variables, any F satisfying

$$\int_{p_i^s,\theta_i\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta_i}(p_i^s,\theta_i,\zeta_i) = \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF^*_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i), \tag{6}$$

also yields the same value for deadweight loss:

$$\int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) = DWL.$$

The main message of this section will be the failure of such a result obtaining. We show that there are at least two distributions satisfying 6 that yield different values of DWL. These two distributions also turn out to yield tight bounds to the possible values of DWL that are consistent with aggregate demand.

Finally, we impose regularity conditions throughout this section to rule out ill-defined integrals. Formally, we insist that the econometrician only consider distributions that satisfy the *integrability conditions*, described below.

Definition 1. *A distribution $F_{p^s,\theta,\zeta}$ satisfies the integrability conditions if:*

1. *q and dwl are integrable on any measurable set.*

2. *$q(p;\theta_i,z)$ is integrable on any subset of the support of θ for any $p>0$ and any z in the range of ζ.*

For instance, all distributions with a finite support of (p^s,θ,ζ) satisfy the above conditions.

3.1 Lower Bound on Deadweight Loss

Consider arbitrary $\bar{p}$, p^{NT}, and τ. For arbitrary $F_{p^s,\theta,\zeta}$ consistent with the data, we can choose a price $\hat{p}^s$ that could also rationalize the data if perceived homogeneously.

Proposition 1. *For any $F_{p^s,\theta,\zeta}$ that yields integrable aggregate demand, there exists $\hat{p}^s$ such that for some distribution $F'_{\theta,\zeta}$ such that $F'_\theta = F_\theta$:*

$$\int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) = \int_{\theta_i,\zeta_i} q(\hat{p}^s;\theta_i,\zeta_i)dF'_{\theta,\zeta}(\theta_i,\zeta_i).$$

We can always rationalize the data with a joint distribution of (p^s,θ) in which θ has marginal distribution F^*_θ, whereas $p^s=\hat{p}^s$ with probability one. We now show that such a joint distribution provides a generic underestimate to the possible values of deadweight loss.

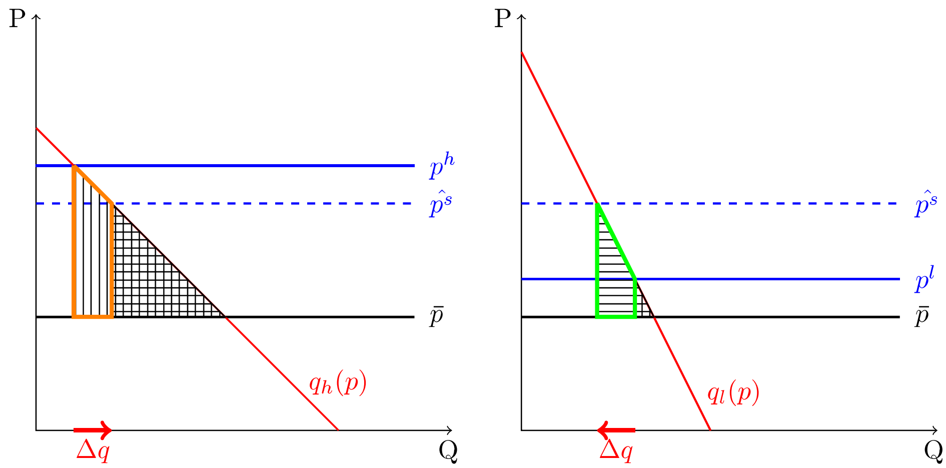
Theorem 1. *Consider any joint distributions $F_{p^s,\theta,\zeta}$ and $F_{\theta,\zeta}$ with corresponding value $\hat{p}^s$ such that:*

$$\int_{\theta,\zeta} q(\hat{p}^s;\theta_i,\zeta_i)dF_{\theta,\zeta}(\theta_i,\zeta_i) = \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).$$

Then the following inequality obtains:

$$\int_{\theta_i,\zeta_i} dwl(\hat{p}^s;\theta_i,\zeta_i)dF_{\theta,\zeta}(\theta_i,\zeta_i) \leq \int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).$$

Intuitively, introducing heterogeneity in perceived prices can mute gains from trade. If someone with a higher marginal valuation for the good has a higher perceived price than someone with a lower marginal valuation, they could both gain by trading with each other after making their consumption decisions. If they could exchange with each other, the one who perceived the higher price could purchase some of the good from the other agent, making both agents better off. Thus, ruling out perceived price heterogeneity by assuming a homogeneous perceived price $\hat{p}^s$ eliminates the possibility of an allocative inefficiency. Figure 3 offers graphical intuition for the lower bound.



(a) High perceived price

(b) Low perceived price

Figure 3: A graphical illustration of Theorem 1. When one picks $\hat{p}^s$ as to make the change in demand equal for the consumer in (a) and in (b), the decrease in dwl for the consumer in (a) (orange) must be at least as large as the increase in dwl for the consumer in (b) (green).

Theorem 1 points out that for any distribution that rationalizes the data, i.e. that explains the observed aggregate demand, one can alternatively rationalize the data with a homogeneous perceived price that yields (weakly) less deadweight loss. From this, we can reach two conclusions. One, we generally cannot identify deadweight loss because we could always alternatively rationalize the data with a homogeneous perceived price.[11] This holds even if we already knew the distribution of preference types F_θ^*. Two, if there is a minimum value of deadweight loss that is consistent with the data, that value of deadweight loss comes from a distribution with no heterogeneity in tax salience.

3.2 Upper bound on deadweight loss

The upper bound comes from an assumption on the limits to tax salience:

Assumption 1. *There is some value $\bar{m} \geq 0$ such that p^s has support known to be contained in $\mathcal{P} \equiv [\bar{p}, \bar{p}+\bar{m}\tau]$.*

[11] This claim holds generically, but would not hold, for instance, if there was no heterogeneity in tax salience.

This assumption says that agents must perceive a non-negative tax τ^s no greater than fraction $\bar{m}$ of the true tax.[12] The econometrician is allowed to assume an arbitrarily large $\bar{m}$, but the gain in robustness will likely come at the expense of precision. For instance, setting $\bar{m} = 1$ would be to assume that agents never over-react to a tax rate. Imposing that $\tau^s \geq 0$ with probability one already ensures that deadweight loss is no greater than the original consumer surplus.[13] But the interval restriction implies any distribution yields no more deadweight loss than a distribution with "binary" perceived prices, i.e. where p^s can only take on values in $\{\bar{p}, \bar{p} + \bar{m}\tau\} \equiv \partial\mathcal{P}$.

The gist of the upper bound of deadweight loss is that, for any data-generating process that rationalizes observed aggregate demand, there is another data-generating process that also rationalizes observed demand, but which would yield at least as much deadweight loss. This alternative explanation of the observed demand insists that all agents pay either zero or maximal attention.

Before formally stating our main result, we demonstrate how one can always pick such a distribution of attention to rationalize observed demand, for any underlying (known) distribution of preferences. Then, we state the main result, theorem 2, providing some intuition for why such a distribution would yield a weakly higher deadweight loss. Because our model of choice may result in several choices given the same sticker prices and taxes, sometimes there can be multiple equilibria with the same level of aggregate demand, but different consequences for welfare. We briefly discuss appendix theorem 3, which deals deals with such cases.

Consider any $F_{p^s,\theta,\zeta}$ that rationalizes the data, and such that:

$$\lim_{m \to \bar{m}^-} F_{p^s}(\bar{p} + m\tau) - F_{p^s}(\bar{p}) > 0.$$

In words, the distribution assumes some positive mass of agents pay neither zero nor maximal attention, i.e. $m \in (\bar{p}, \bar{p} + \bar{m}\tau) \equiv int(\mathcal{P})$. Pick $\tilde{p}^s \in int(\mathcal{P})$ and a corresponding $p^b(p_i^s) \equiv \bar{p} + \mathbb{I}(p_i^s > \tilde{p}^s)\bar{m}\tau$ such that:

$$\begin{aligned}\int_{p_i^s \in int(\mathcal{P}),\theta_i} q(p^b(p_i^s);\theta_i, l)dF_{p^s,\theta}(p_i^s,\theta_i) &\leq \int_{p_i^s \in int(\mathcal{P}),\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \\ &\leq \int_{p_i^s \in int(\mathcal{P}),\theta_i} q(p^b(p_i^s);\theta_i, h)dF_{p^s,\theta}(p_i^s,\theta_i).\end{aligned}$$

In words, for any distribution that puts mass on $int(\mathcal{P})$, we pick a value $\tilde{p}^s$ that acts as a divide: those below it get assigned to a group that does not perceive the tax at all, while those above it get assigned to a group that perceives it "maximally". Since demand is monotonic in p, and given our definitions of l and h, one can always pick $\tilde{p}^s$ such that the above inequalities hold weakly. Thus, one can always find $\lambda \in [0, 1]$ such that:

$$\begin{aligned}\lambda \int_{p_i^s \in int\mathcal{P},\theta_i} q(p^b(p_i^s);\theta_i, h)dF_{p^s,\theta}(p_i^s,\theta_i) + (1-\lambda)\int_{p_i^s \in int\mathcal{P},\theta_i} q(p^b(p_i^s);\theta_i, l)dF_{p^s,\theta}(p_i^s,\theta_i) \\ = \int_{p_i^s \in int\mathcal{P},\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).\end{aligned} \tag{7}$$

Equation 7 implies that we can always pick a threshold $\tilde{p}^s$ that rationalizes demand, so long as we randomly assign a fraction λ of consumers with $m \in int(\mathcal{P})$ to the tie-breaking parameter $\zeta = h$, and the remaining $1 - \lambda$ to $\zeta = l$. But in turn, this implies that we have found binary distribution of p^s and ζ that rationalizes aggregate demand.

Let us call such distribution $F''_{p^s,\theta,\zeta}$. This new distribution has the same marginal distribution of preferences, $F_\theta = F''_\theta$. If a consumer perceived a price in the boundary region in the original distribution $F_{p^s,\theta,\zeta}$, then so will she in the new distribution: $F''_{p^s,\theta,\zeta|p^s\in\partial\mathcal{P}} = F_{p^s,\theta,\zeta|p^s\in\partial\mathcal{P}}$. As for those consumers who perceived a price in the interior, we propose to split them up in a manner akin to what we just did in equation 7: $F''_{p^b(p^s),\theta} = F_{p^s,\theta}$, and ζ is assigned randomly as we described above. One can quickly confirm that this

[12] This description implicitly assumes that $\tau > 0$.

[13] Recall that deadweight loss equals its calculation as if taxes actually satisfied $\tau_i = p_i^s - \bar{p}$, so excess burden cannot exceed the original consumer surplus for any agent. One can show that if τ^s has support on negative values, then it's possible to have total deadweight loss substantially greater than the original total consumer surplus.

distribution rationalizes the same demand as the original distribution $F_{p^s,\theta,\zeta}$:

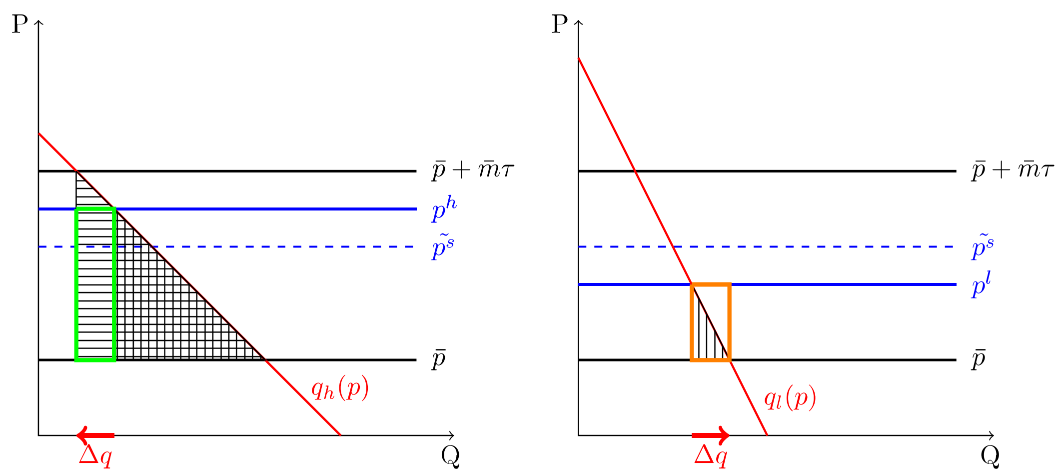
$$\begin{aligned}\int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) &= \int_{p_i^s\in int(\mathcal{P}),\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &+ \int_{p_i^s\in\partial\mathcal{P},\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &= \int_{p_i^s\in int(\mathcal{P}),\theta_i} [\lambda q(p^b(p_i^s);\theta_i,h) + (1-\lambda)q(p^b(p_i^s),\theta_i,l)]dF_{p^s,\theta}(p_i^s,\theta_i)\\ &+ \int_{p_i^s\in\partial\mathcal{P},\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &= \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).\end{aligned}$$

Furthermore, such a distribution provides a generically larger value of deadweight loss than does $F_{p^s,\theta,\zeta}$.

Theorem 2. *Under assumption 1, for any $F_{p^s,\theta,\zeta}$ and any corresponding $F''_{p^s,\theta,\zeta}$ as described above:*

$$\int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \geq \int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i).$$

We can obtain intuition in two ways. One is to note that the method of forcing binary perceived prices increases heterogeneity of perceived prices compared to $F_{p^s,\theta,\zeta}$. Another is by considering the case where $\bar{m} = 1$ and F_θ^* is known to be degenerate, so that all agents have the same preferences. For a given aggregate demand, deadweight loss is maximized under these preferences when some perceive price $p_i^s = \bar{p}$, while others correctly perceived the true tax rate $p_i^s = \bar{p} + \tau$. This is because for each individual agent, deadweight loss is convex in the perceived price. Hence, for a given aggregate demand, *aggregate* deadweight loss will be highest when it is as high as possible for some – namely, those who fully perceive the tax – while it is null for everybody else – as those who don't perceive the tax at all are effectively subject to a lump-sum tax. We provide a graphical illustration of this argument in figure 4.



(a) High perceived price

(b) Low perceived price

Figure 4: A graphical illustration of Theorem 2. The watershed price $\tilde{p}^s$ is chosen to make the change in demand equal for the consumer in (a) and in (b). As long as we are dealing with weakly decreasing demand functions, the increase in deadweight loss for (a) is at least as big as the green box, while the decrease in deadweight loss for (b) is at most as big as the orange box. By assigning a perceived price of $\bar{p} + \bar{m}\tau$ to the consumer in (a) and $\bar{p}$ to the consumer in (b), we have increased aggregate deadweight loss holding aggregate demand constant.

Theorem 2 illustrates that for any distribution of (p^s, θ) that rationalizes the data, we can alternatively rationalize the data with a distribution with support for p_i^s on $\{\bar{p}, \bar{p} + \bar{m}\tau\}$ that yields (weakly) greater deadweight loss. Again, we see that identification of deadweight loss is not generally possible even if we knew the distribution of F_θ^*, as different marginal distributions of p^s and ζ could have different implications for deadweight loss. Also, any upper bound to the possible values of deadweight loss must be generated from a distribution with support of perceived prices on $\{\bar{p}, \bar{p} + \bar{m}\tau\}$.

However, not all distributions that have $p^s \in \partial\mathcal{P}$ with probability one yield the same value of deadweight loss, even when rationalizing the same data with the same distribution of preference types. The allocation of the good that yields the highest possible deadweight loss will also assign more consumption to agents with more convex demand curves. Theorem 3 in the online appendix spells out how to assign consumption of the good in the way it will "do the least good", and deals with cases where the tie-breaking parameter ζ is relevant, to get a general expression for the upper bound for deadweight loss consistent with aggregate demand and the distribution of preferences.

4 Linear Special Case

In this section, we discuss the special case in which q is known to be linear in $\bar{p}$ and τ (for fixed p^{NT}). We focus on this example both because of how frequently economists estimate linear models and because of its relationship to the second order approximation of deadweight loss.

We can also use the linear special case to better illustrate the general identification problem. In this subsection, we will no longer assume that the distribution of preference parameters is known; the econometrician will, as is usually the case, be able to identify preferences with exogenous price variation. As in TRJ (2018), we assume the distribution of preferences does not depend on taxes (or prices). Further, we permit the econometrician to assume that the distribution of tax salience does not change as sticker prices and taxes vary. This entirely rules out any sort of endogeneity between attention and taxes – which was driving non-identification in the binary case – and yet we will get the same non-identification result.

Naturally, no demand curve can be entirely linear, for the simple reason that agents cannot consume negative quantities. But in practice, one rarely gets the privilege of such rich variation in prices when doing empirical work. What we are implicitly assuming in this section is that the values of $(\bar{p}, \tau)$ that are considered are all such that $\bar{p} > 0$, and $q_i > 0$ for all consumers in the market, so that aggregate demand is also linear at those values. In other words, one can think of our work here as modeling linear demand conditional on buying at all prices under consideration.[14]

We might recall from section 2 that one can express a second order approximation to deadweight loss as a function of derivatives. If the choice function is linear in regressors $\bar{p}$ and τ, the second order approximation is an exact calculation of deadweight loss, and our results from the previous sections apply.

Formally, each preference type θ_i takes the form $\theta_i = (\beta_i, \epsilon_i) \in \mathbb{R}^2$.[15] To maintain linearity in regressors, we also assume that tax salience m is constant with respect to τ. The choice function q then takes the form:

$$q_i = \alpha + \beta_i p_i^s + \epsilon_i = \alpha + \beta_i[\bar{p} + m_i\tau] + \epsilon_i$$

We are suppressing the tie-breaking parameter ζ because in this linear example $\theta_i, m_i, \bar{p}$, and τ always uniquely determine consumption. We have the parameter α so that we can assume without loss of generality that $\mathbb{E}[\epsilon] = 0$.

Defining $\tilde{\beta}_i \equiv m_i\beta_i$ yields:

$$q_i = \alpha + \beta_i\bar{p} + \tilde{\beta}_i\tau + \epsilon_i, \tag{8}$$

[14] We thank an anonymous referee at the Journal of Public Economic Theory for helping us clarify our thinking on this matter.

[15] Agents have quasi-linear utility $u_i = \frac{q_i^2/2 - (\alpha+\epsilon_i)q_i}{\beth_i} + q_i^{NT}$. For a given p^{NT}, we define $\beta_i \equiv \frac{\beth_i}{p^{NT}}$, yielding utility representation $U_i = \frac{q_i^2/2 - (\alpha+\epsilon_i)q_i}{\beta_i} + p^{NT}q_i^{NT}$.

with corresponding deadweight loss per agent, from equation 4:

$$\begin{aligned}
dwl_i &= \int_{\bar{p}}^{p^s} [\alpha + \beta_i p + \epsilon_i]dp - (p^s - \bar{p})[\alpha + \beta_i p^s + \epsilon_i] = \int_{\bar{p}}^{p^s} (p - p^s)\beta_i dp \\
&= \frac{1}{2}\Big[\frac{{p^s}^2 - \bar{p}^2}{2} - (p^s - \bar{p})p^s\Big]\beta_i = \frac{1}{2}(p^s - \bar{p})[(p^s + \bar{p}) - 2p^s]\beta_i = -\frac{1}{2}{\tau^s}^2\beta_i \\
&= -\frac{1}{2}m_i^2\beta_i\tau^2.
\end{aligned}$$

We assume that not just the distribution of preference parameters, but also the joint distribution of preference and salience parameters remains unaffected by the specific values of $\bar{p}$ and τ. The econometrician observes for various values of regressors:

$$\mathbb{E}[q|\bar{p},\tau] \equiv \int\limits_{\beta_i,\tilde{\beta}_i,\epsilon_i} [\alpha + \beta_i\bar{p} + \tilde{\beta}_i\tau + \epsilon_i]dF^*_{\beta,\tilde{\beta},\epsilon}(\beta_i,\tilde{\beta}_i,\epsilon_i) = \alpha + \mathbb{E}[\beta]\bar{p} + \mathbb{E}[\tilde{\beta}]\tau. \tag{9}$$

where $F^*_{\beta,\tilde{\beta},\epsilon}$ is the true distribution of $(\beta, m\beta, \epsilon)$. The challenge is to use the observed values of triplets $(\bar{p}, \tau,\ \mathbb{E}[q|\bar{p},\tau])$ to infer aggregate deadweight loss, which in this case is equivalent to its second order approximation around $\tau = 0$:

$$DWL = -\frac{1}{2}\int_{\beta_i,m_i} m_i^2\beta_i dF_{\beta,m}(\beta_i,m_i)\tau^2 = -\frac{1}{2}\mathbb{E}[m^2\beta]\tau^2 = -\frac{1}{2}\mathbb{E}[m\tilde{\beta}]\tau^2.$$

The only restriction that the econometrician imposes on the distribution of tax salience m is that the support of tax salience is contained within the interval $[0, \bar{m}]$. The econometrician can also use the Compensated Law of Demand as defined in appendix lemma 2, which shows that compensated demand is always weakly decreasing, so that $\mathbb{P}[\beta \le 0] = 1$. In fact, we can permit the econometrician to know the entire distribution of $\theta = (\beta, \epsilon)$. It will not affect our results.} First, we can find a homogeneous perceived price that rationalizes the data for any τ. In particular, a linear regression of aggregate demand on sticker prices and taxes may permit identification of $\hat{\beta} \equiv \mathbb{E}[\beta]$ and $\hat{\tilde{\beta}} \equiv \mathbb{E}[\tilde{\beta}]$, respectively.[16] We define a measure of central tendency of tax salience:[17]

$$\hat{m} \equiv \frac{\hat{\tilde{\beta}}}{\hat{\beta}}.$$

Then the homogeneous perceived price that rationalizes the data is $\hat{p}^s = \bar{p} + \hat{m}\tau$. To see this, note that assuming all agents have tax salience $m_i = \hat{m}$ yields aggregate demand as in equation 9:

$$\begin{aligned}
\int_{\beta_i,\epsilon_i} [\alpha + \beta_i\hat{p}^s + \epsilon_i]dF^*_{\beta,\epsilon}(\beta_i,\epsilon_i) &= \alpha + \bar{p}\int_{\beta_i,\epsilon_i} \beta_i dF^*_{\beta}(\beta_i) + \hat{m}\tau\int_{\beta_i} \beta_i dF^*_{\beta}(\beta_i) \\
&= \alpha + \hat{\beta}\bar{p} + \hat{m}\hat{\beta}\tau \\
&= \alpha + \hat{\beta}\ \bar{p} + \hat{\tilde{\beta}}\ \tau.
\end{aligned}$$

Thus, the econometrician cannot rule out all agents perceiving the same price $\hat{p}^s$, and so cannot rule out $m_i = \hat{m}\ \forall i$. For tax τ, this would yield deadweight loss:

$$DWL_{low} = -\frac{1}{2}\hat{m}\hat{\tilde{\beta}}\tau^2.$$

By theorem 1, this is a lower bound for deadweight loss.

Alternatively, the econometrician cannot rule out the perceived tax τ^s having support in $\{0, \bar{m}\tau\}$. To

[16]Such identification requires exogenous and non-collinear variation in sticker prices and taxes. If the econometrician cannot identify these terms, so much the worse for identifying aggregate deadweight loss.

[17]If $\hat{\beta} = 0$, then let $\hat{m} = 0$.

see this, consider $\mathbb{P}(p^s = \bar{p} + \bar{m}\tau) = \frac{\hat{m}}{\bar{m}}$ and $\mathbb{P}(p^s = \bar{p}) = 1 - \frac{\hat{m}}{\bar{m}}$ independently of other parameters and regressors.[18] This will rationalize aggregate demand:

$$\int_{\beta_i, \epsilon_i} \Big[\alpha + \beta_i \bar{p}_i + \frac{\hat{m}}{\bar{m}} \beta_i \bar{m} \tau_i + \epsilon_i \Big] dF_{\beta,\epsilon}(\beta_i, \tilde{\beta}_i, \epsilon_i) = \alpha + \mathbb{E}[\beta]\bar{p} + \hat{m}\mathbb{E}[\beta]\tau = \alpha + \mathbb{E}[\beta]\bar{p} + \mathbb{E}[\tilde{\beta}]\tau.$$

This yields deadweight loss for tax τ:

$$DWL_{high} = -\frac{1}{2}\frac{\hat{m}}{\bar{m}}\mathbb{E}[\beta]\bar{m}^2\tau^2 = -\frac{1}{2}\hat{m}\hat{\beta}\bar{m}\tau^2 = -\frac{1}{2}\bar{m}\hat{\tilde{\beta}}\tau^2.$$

For instance, if $\bar{m} = 1$, then the value of deadweight loss under a homogeneous perceived price is fraction $\hat{m}$ of the above calculation of deadweight loss.

Proceeding from theorem 2, we noted that there is a specific distribution of perceived prices on $\{\bar{p}, \bar{p}+\bar{m}\tau\}$ that maximizes deadweight loss. We describe that distribution in theorem 3, noting that it involves assigning high or low perceived prices based on the ratio of per-person deadweight loss to the change in consumption for that individual. But in this context:

$$\frac{dwl_i}{q_i(\bar{p}) - q_i(p^s)} = \frac{\tau^s}{2}.$$

Thus, the distribution of tax salience independent of all other parameters and regressors in which $\mathbb{P}(m = \bar{m}) = \frac{\hat{m}}{\bar{m}}$ and $\mathbb{P}(m = 0) = 1 - \frac{\hat{m}}{\bar{m}}$ maximizes deadweight loss. More generally, the econometrician cannot rule out this maximal value of deadweight loss so long as they cannot rule out the possibility of some distribution F with $F_\beta = F^*_\beta$ such that $supp(m) \in \{0, \bar{m}\}$ with:

$$\mathbb{P}_F(m = \bar{m})\mathbb{E}_F[\tilde{\beta}|m = \bar{m}] = \hat{m}\hat{\beta} = \hat{\tilde{\beta}}.$$

One can check that this distribution rationalizes the data,

$$\mathbb{E}[q|\bar{p}, \tau] = \alpha + \mathbb{E}[\beta]\bar{p} + \mathbb{E}_F[\tilde{\beta}]\tau = \alpha + \hat{\beta}\bar{p} + \mathbb{P}_F[m = \bar{m}]\mathbb{E}_F[\tilde{\beta}|m = \bar{m}]\tau = \alpha + \hat{\beta}\bar{p} + \hat{\tilde{\beta}}\tau,$$

and yields the maximal value of deadweight loss,

$$-\frac{1}{2}\mathbb{E}_F[m^2\beta]\tau^2 = -\frac{1}{2}\mathbb{P}_F[m = \bar{m}]\bar{m}\mathbb{E}_F[\tilde{\beta}|m = \bar{m}]\tau^2 = -\frac{1}{2}\bar{m}\hat{\tilde{\beta}}\tau^2 = DWL_{high}.$$

More intuitively, once one knows $\hat{\beta}$ and $\hat{\tilde{\beta}}$, one can rationalize the aggregate data. Since the ratio of deadweight loss to the change in quantity is constant, the relationship between tax salience and preferences doesn't matter upon attaining the observed aggregate demand.

Finally, consider a distribution with $m \perp (\beta, \epsilon)$ with $supp(m) \subseteq \{0, \hat{m}, \bar{m}\}$, $\mathbb{P}(m = \hat{m}) = \lambda$ and $\mathbb{P}(m = \bar{m}|m \neq \hat{m}) = \frac{\hat{m}}{\bar{m}}$. Varying λ from zero to one yields:

$$DWL \in \Big[-\frac{1}{2}\hat{m}\hat{\tilde{\beta}}\tau^2, -\frac{1}{2}\bar{m}\hat{\tilde{\beta}}\tau^2\Big].$$

We can conclude from this result that one cannot even identify a second order approximation of deadweight loss with aggregate data alone.[19] Imposing structure on preferences to facilitate identification of F^*_θ still only permits interval identification. Nonetheless, we can use aggregate data to obtain bounds, or at least $\hat{m}$, which gives us a sense of the uncertainty over the possible values of deadweight loss.

Besides its illustrative value, this linear framework also gives researchers a quick and easy way to compute bounds for the deadweight loss of non-salient taxes or fees in a variety of empirical contexts. In the online appendix, we apply these findings to the framework of CLK's study of aggregate beer consumption and Goldin and Homonoff's (2013) study of cigarette consumption. Details on the estimation procedures are

[18] In the true distribution, it must be that $\hat{m} \in [0, \bar{m}]$. Alternatively, one could consider checking whether $\hat{m} \in [0, \bar{m}]$ as a weak test of the null hypothesis that tax salience is bounded within that interval.

[19] One can identify a first order approximation trivially; it is zero.

provided in the online appendix. In the baseline specification of the CLK (2009) data, we estimate that $\hat{m} \approx 0.31$. This estimate suggests that even assuming that salience cannot exceed one, $\bar{m} = 1$, the upper bound of deadweight loss is about three times the lower bound. These estimates, however, all seem fairly imprecise. Across the two data-sets, there is no specification in which we can reject the null hypothesis that $\hat{m} = 0$, permitting the upper bound to be arbitrarily large in proportion to the lower bound.[20] Similarly, in most specifications we cannot reject that $\hat{m} = 1$, which would imply that upper bound and lower bound are identical. This underlying uncertainty is mirrored in previous work on tax salience – e.g., TRJ (2018) find that individual differences increase excess burden by at least 200% relative to the case of homogeneous attention – but our wide confidence intervals may be the result of the specific data sets we are using here – for example, Goldin and Homonoff (2013) often cannot reject that consumers do not react at all to sales taxes. Our procedure, however, seems so straightforward to carry out that it might turn out useful in future research on tax salience.

Finally, in both the setting of CLK (2009) and Goldin and Homonoff (2013), functional form assumptions seem to matter. The limitations of the linear setting would prompt us to undergo more sophisticated and less parametric exercises, but we are dissuaded by the fact that our statistical power is already very low.

5 Conclusion

In this paper, we studied deadweight loss in a model where agents misperceive prices. We started by generalizing the theoretical results of CLK (2007, 2009) with an arbitrary closed choice set. Inattentiveness to taxes makes agents worse off while reducing deadweight loss by preventing agents from avoiding the tax.

As in the binary choice model of TRJ (2018), heterogeneous attention adds another layer of complexity to deadweight loss. In our general setting, we show that aggregate consumption can be consistent with a wide variety of co-distributions of attention and preferences, each with a different implication for deadweight loss. This is because it matters who gets the good and who doesn't: when prices are misperceived, there is no guarantee that agents who end up with some units of the good are the ones who value those units most. By minimizing and maximizing this allocative inefficiency, we show that deadweight loss can only vary between two extremes for any given aggregate demand. The lower bound holds generally, while the upper bound relies on the assumption that tax salience has support contained in a known non-negative interval.

Finally, we explore the special case in which demand is linear, which is of special interest due to its relationship to both the empirical literature and the second order approximation of deadweight loss. Our analysis shows that, while identification of deadweight loss under binary choice may be restated as an endogeneity problem, the same cannot be said regardless of the choice set. Indeed, when individual demand is linear, assuming independence of tax salience from taxes and prices does not change the interval of possible values of deadweight loss.

The linear model yields bounds for deadweight loss that one can easily compute from linear regression estimates. While this doesn't necessarily doom any future application in empirical work, our own applications of this method on the existing work of CLK (2009) and Goldin and Homonoff (2013) leave us without many answers about how tight these bounds might be. While some point estimates seem encouraging, they also can be imprecise and dependent on functional form specification.

[20] Note that $\frac{DWL_{high}}{DWL_{low}} = \frac{\bar{m}}{\hat{m}}$, so that as $\hat{m}$ approaches zero from above, this ratio of upper to lower bounds blows up to infinity.

References

Hunt Allcott and Dmitry Taubinsky. Evaluating behaviorally motivated policy: Experimental evidence from the lightbulb market. *American Economic Review*, 105(8):2501–38, 2015.

B Douglas Bernheim and Antonio Rangel. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. *The Quarterly Journal of Economics*, 124(1):51–104, 2009.

Tom Blake, Sarah Moshary, and Steve Tadelis. Price salience and product choice. *NBER Working Paper N. 25186*, 2017.

Robin Boadway, Jean-Denis Garon, and Louis Perrault. Optimal mixed taxation, credit constraints, and the timing of income tax reporting. *Journal of Public Economic Theory*, 21:708–737, 2018.

Steven Bradley and Naomi Feldman. Hidden baggage: behavioral responses to changes in airline ticket tax disclosure. *Finance and Economics Discussion Series 2018-057. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2018.057*, 2018.

Raj Chetty. The simple economics of salience and taxation. *NBER Working Paper N. 15246*, 2009.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation. *NBER Working Paper N. 13330*, 2007.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–1177, 2009.

Amy Finkelstein. Ez-tax: Tax salience and tax rates. *Quarterly Journal of Economics*, 124(3):969–1010, 2009.

Jeremy T Fox. A note on nonparametric identification of distributions of random coefficients in multinomial choice models. *NBER Working Paper No. 23621*, 2017.

Xavier Gabaix. A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, 129(4): 1661–1710, 2014.

Kelly Sims Gallagher and Erich Muehlegger. Giving green to get green? incentives and consumer adoption of hybrid vehicle technology. *Journal of Environmental Economics and Management*, 61(1):1–15, 2011.

Jacob Goldin. Optimal tax salience. *Journal of Public Economics*, 131:115–123, 2015.

Jacob Goldin and Tatiana Homonoff. Smoke gets in your eyes: cigarette tax salience and regressivity. *American Economic Journal: Economic Policy*, 5(1):302–336, 2013.

Matthew A Masten. Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies*, 85(2):1193–1250, 2017.

Andrea Morone, Francesco Nemore, and Simone Nuzzo. Experimental evidence on tax salience and tax incidence. *Journal of Public Economic Theory*, 20(4):582–612, 2018.

Harvey S Rosen. Tax illusion and the labor supply of married women. *The Review of Economics and Statistics*, 58(2):167–172, 1976.

Dmitry Taubinsky and Alex Rees-Jones. Attention variation and welfare: theory and evidence from a tax salience experiment. *Review of Economic Studies*, 85(4):2461–2496, 2018.

Matthias Weber and Arthur Schram. The non-equivalence of labour market taxes: A real-effort experiment. *The Economic Journal*, 127(604), 2016.

ONLINE APPENDIX FOR:
Attending to Inattention: Identification of Deadweight Loss under Non-Salient Taxes

Giacomo Brusco
Economics, University of Michigan at Ann Arbor

Benjamin Glass
Economics, Pennsylvania State University at University Park

A Additional Results and Proofs

A.1 Additional results and proofs from section 2

Proposition 2. *Suppose we can extend u to $\mathbb{R}^2$ such that u is continuous and quasi-concave. Then for any $\bar{p}$, p^{NT}, τ, and W on which $\boldsymbol{q}$ is defined, there exist scalar values p^s and W^s such that:*

$$\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W) \in \underset{p^s q + p^{NT} q^{NT} \leq W^s}{\arg\max} u(q, q^{NT})$$
$$\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W^s) * (\bar{p} + \tau, p^{NT}) = W$$

We demonstrate a generalization of proposition 2, in which multiple goods may be taxed. We consider a general setting with N goods, consumption set $X = X^T \times X^{NT} \subseteq \mathbb{R}^N_+$, with consumption vector $\boldsymbol{q} = (\boldsymbol{q^T}, \boldsymbol{q^{NT}}) \in X$. Here X^T is the consumption set for taxed goods, while X^{NT} is the consumption set for non-taxed goods. We assume that either $X^{NT} \subseteq \mathbb{R}_+$ or X^{NT} is convex.

The agent has preferences $\succeq$ on X. Informally, we want to assume preferences such that agents smoothly prefer moderation. To say that they prefer moderation, one generally assumes convex preferences. However, we do not want to assume a convex consumption set X. We might alternatively assume that preferences are *pseudo-convex*, in that for any $\boldsymbol{q} \in X$ and any finite n:

$$\boldsymbol{q}_k \in X, \boldsymbol{q}_k \succ \boldsymbol{q}, \lambda_k \geq 0 \ \forall k = 1, ..., n, \sum_{k=1}^{n} \lambda_k = 1, \sum_{k=1}^{n} \lambda_k \boldsymbol{q}_k \in X \Rightarrow \sum_{k=1}^{n} \lambda_k \boldsymbol{q}_k \succ \boldsymbol{q}$$

However, we also want some smoothness to preferences. More formally, we want to figure that if $\boldsymbol{q'} \succ \boldsymbol{q}$, then there is an epsilon ball around $\boldsymbol{q'}$ such that the agent would prefer any element in that epsilon ball to $\boldsymbol{q}$ if that element were also in the consumption set. Furthermore, any convex combination of points in these epsilon balls should yield a point that, if contained in X, is also strictly preferred to $\boldsymbol{q}$. We refer to this assumption on preferences as *continuous pseudo-convexity* (CPC).

Assumption 2. *For any $\boldsymbol{q} \in X$, define the set of strictly preferred allocations:*

$$\mathcal{A} \equiv \{\boldsymbol{q'} \in X | \boldsymbol{q'} \succ \boldsymbol{q}\}$$

There exists some function $\epsilon : \mathcal{A} \rightarrow \mathbb{R}_{++}$ such that for any $n \in \mathbb{N}$, for any $\lambda_1, ..., \lambda_n \geq 0$ and $\boldsymbol{q}_1, ..., \boldsymbol{q}_n \in \mathcal{A}$, if $\sum_{k=1}^{n} \lambda_k = 1$, then :

$$\exists \boldsymbol{q}'_1, ..., \boldsymbol{q}'_n \in \mathbb{R}^n : ||\boldsymbol{q}'_k - \boldsymbol{q}_k|| < \epsilon(\boldsymbol{q_k}) \ \forall k, \sum_{k=1}^{n} \lambda_k \boldsymbol{q_k}' \in X \Rightarrow \sum_{k=1}^{n} \lambda_k \boldsymbol{q_k}' \succ \boldsymbol{q}$$

We provide this description of CPC preferences to facilitate intuition, but our main result for this section comes from an equivalent, yet more geometric, expression of this description.

Lemma 1. *Preferences $\succeq$ are CPC if and only if for every $\boldsymbol{q} \in X$ with corresponding set of strictly preferred bundles $\mathcal{A}$ there is an open and convex set $\mathcal{O} \subseteq \mathbb{R}^N$ such that $\mathcal{O} \cap X = \mathcal{A}$.*[1]

Proof: For one direction, the convex hull of the union of open $\epsilon(\boldsymbol{q}')$ balls around $\boldsymbol{q}' \in \mathcal{A}$ is open, and by assumption does not contain any elements of $X \setminus \mathcal{A}$. For the other direction, for any $\boldsymbol{q}' \in \mathcal{A}$, define $\epsilon(\boldsymbol{q}')$ as a positive value such that $\boldsymbol{q}'' \in \mathbb{R}^N_+ : ||\boldsymbol{q}'' - \boldsymbol{q}'|| < \epsilon(\boldsymbol{q}') \Rightarrow \boldsymbol{q}'' \in \mathcal{O}$. We can do so because $\mathcal{O}$ is open. For any such $\boldsymbol{q}''$, if $\boldsymbol{q}'' \in X$, then $\boldsymbol{q}'' \succ \boldsymbol{q}$. $\square$

Let $\boldsymbol{p} = (\boldsymbol{p^T}, \boldsymbol{p^{NT}}) \in \mathbb{R}^N_+$ denote a generic price vector, where $\boldsymbol{p^T}$ and $\boldsymbol{p^{NT}}$ are price vectors for taxed and non-taxed goods respectively. In particular, let $\bar{\boldsymbol{p}} = (\bar{\boldsymbol{p}}^{\boldsymbol{T}}, \bar{\boldsymbol{p}}^{\boldsymbol{NT}})$ denote the vector of sticker prices.

Let $\boldsymbol{\tau}$ denote the vector of taxes for taxed goods, so that $\boldsymbol{q^T}$, $\boldsymbol{q^{NT}}$, and $\boldsymbol{\tau}$ all have the same number of elements. The consumption vector $\boldsymbol{q}(\bar{\boldsymbol{p}}, \boldsymbol{\tau}) = (\boldsymbol{q}^T(\bar{\boldsymbol{p}}, \boldsymbol{\tau}), \boldsymbol{q}^{NT}(\bar{\boldsymbol{p}}, \boldsymbol{\tau}))$ satisfies the following properties:

$$\begin{aligned}
\bar{p}^{NT} * \tilde{\boldsymbol{q}}^{NT} &\leq W - \bar{\boldsymbol{p}}^{\boldsymbol{T}} * \boldsymbol{q}^T \\
(\boldsymbol{q}^T, \tilde{\boldsymbol{q}}^{NT}) &\succ (\boldsymbol{q}^T, \hat{\boldsymbol{q}}^{NT}) \;\; \forall \hat{\boldsymbol{q}}^{NT} \in X^{NT} : \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \tilde{\boldsymbol{q}}^{NT} \leq W - \bar{\boldsymbol{p}}^{\boldsymbol{T}} * \boldsymbol{q}^T \\
\boldsymbol{q}(\bar{\boldsymbol{p}}, \boldsymbol{0}) &\in \underset{\tilde{\boldsymbol{q}} \in X : \bar{\boldsymbol{p}} * \tilde{\boldsymbol{q}} \leq W}{\arg\max} \succeq
\end{aligned}$$

In words, consumption of the non-taxed goods is always optimally determined upon choosing consumption of the taxed goods, and consumption is optimally determined when the agent correctly perceives prices, i.e. when there are no taxes. We also restrict the domain of sticker prices and taxes so that expenditure on non-taxed goods is positive, i.e.:

$$\bar{\boldsymbol{p}}^{NT} * \boldsymbol{q}^{NT}(\bar{\boldsymbol{p}}, \boldsymbol{\tau}) > 0$$

The claim is that for any $\bar{\boldsymbol{p}}$ and $\boldsymbol{\tau}$ in this domain, there is a (p^s, W^s) that explains $\boldsymbol{q}(\bar{\boldsymbol{p}}, \boldsymbol{\tau})$.

Proof of Generalization of Proposition 2: Define $\boldsymbol{q} = (\boldsymbol{q^T}, \boldsymbol{q^{NT}}) = \boldsymbol{q}(\bar{\boldsymbol{p}}, \boldsymbol{\tau})$ and:

$$\mathcal{A}^e \equiv \{(\boldsymbol{q^{T'}}, e^{NT'}) | \boldsymbol{q^{T'}} \in X^T, \exists \boldsymbol{q^{NT'}} \in X^{NT} : \bar{\boldsymbol{p}}^{NT} * \boldsymbol{q^{NT'}} = e^{NT'}, (\boldsymbol{q}^{T'}, \boldsymbol{q}^{NT'}) \in \mathcal{O}\}$$

Suppose for the sake of contradiction that $(\boldsymbol{q^T}, \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \boldsymbol{q^{NT}}) \in Co(\mathcal{A}^e)$, i.e. that $\exists n \in \mathbb{N}$, $(\boldsymbol{q}_k^T, e_k^{NT}) \in \mathcal{A}^e$, and $\lambda_k \geq 0 \; \forall k = 1, ..., n$ such that $\sum_{k=1}^n \lambda_k = 1$ and:

$$\sum_{k=1}^n \lambda_k (\boldsymbol{q}_k^T, e_k^{NT}) = (\boldsymbol{q^T}, \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \boldsymbol{q^{NT}})$$

Since $(\boldsymbol{q}_k^T, e_k^{NT}) \in \mathcal{A}^e \; \forall k$, that means that:

$$\forall k \; \exists \boldsymbol{q}_k^{NT} : e_k^{NT} = \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \boldsymbol{q}_k^{NT}, \boldsymbol{q}_k \equiv (\boldsymbol{q}_k^T, \boldsymbol{q}_k^{NT}) \Rightarrow \boldsymbol{q}_k \in \mathcal{O}$$

If $X^{NT} \subseteq \mathbb{R}_+$, then $\sum_{k=1}^n \lambda_k \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \boldsymbol{q}_k^{NT} = \bar{\boldsymbol{p}}^{\boldsymbol{NT}} * \boldsymbol{q^{NT}}$ implies that $\sum_{k=1}^n \lambda_k \boldsymbol{q}_k^{NT} = \boldsymbol{q^{NT}}$ because positive non-tax expenditure requires that $\bar{\boldsymbol{p}}^{\boldsymbol{NT}} \neq 0$. In that case:

$$\sum_{k=1}^n \lambda_k \boldsymbol{q}_k = \boldsymbol{q} \Rightarrow\Leftarrow \sum_{k=1}^n \lambda_k \boldsymbol{q}_k \in \mathcal{O}$$

This is a contradiction arising from $\boldsymbol{q} \notin \mathcal{O}$.

If X^{NT} is not a subset of $\mathbb{R}_+$, then X^{NT} is convex. This means $\sum_{k=1}^n \lambda_k \boldsymbol{q}_k \in X$. Pseudo-convexity of preferences implies that:

$$\sum_{k=1}^n \lambda_k \boldsymbol{q}_k \succ \boldsymbol{q}$$

[1] Note that $\mathcal{O}$ is open in $\mathbb{R}^N$.

Yet the weighted average of taxed goods is the desired taxed good consumption bundle, whereas the weighted average of non-taxed goods is affordable:

$$\sum_{k=1}^{n} \lambda_k \boldsymbol{q_k^T} = \boldsymbol{q}^T$$

$$\boldsymbol{\bar{p}^{NT}} * \sum_{k=1}^{n} \boldsymbol{q_k^{NT}} = \sum_{k=1}^{n} e_k^{NT} = \boldsymbol{\bar{p}^{NT}} * \boldsymbol{q}^{NT}$$

Thus, the agent could not have optimally chosen $\boldsymbol{q}^{NT}$, another contradiction. We conclude that $(\boldsymbol{q}^T, \boldsymbol{\bar{p}}^{NT} * \boldsymbol{q}^{NT}) \notin Co(\mathcal{A}^e)$.

Now, we can apply the Separating Hyperplane Theorem to say that there is a vector $(\boldsymbol{p}^{Ts}, 1)$, where $\boldsymbol{p}^{Ts}$ has as many elements as $\boldsymbol{q}^T$, such that:

$$(\boldsymbol{p}^{Ts}, 1) * (\boldsymbol{q}^T, \boldsymbol{\bar{p}}^{NT} * \boldsymbol{q}^{NT}) \leq (\boldsymbol{p}^{Ts}, 1) * (\boldsymbol{q^{T'}}, e^{NT'}) \; \forall (\boldsymbol{q^{T'}}, e^{NT'}) \in Co(\mathcal{A}^e)$$

Defining $\boldsymbol{p^s} \equiv (\boldsymbol{p}^{Ts}, \boldsymbol{\bar{p}}^{NT})$, this implies that for any bundle $\boldsymbol{q'} = (\boldsymbol{q}^{T'}, \boldsymbol{q}^{NT'}) \in \mathcal{O}$:

$$\boldsymbol{p}^s * \boldsymbol{q}^{'} \geq \boldsymbol{p}^s * \boldsymbol{q}$$

Since $\mathcal{O}$ is open, the above expression can never be satisfied with equality. To see this, suppose otherwise, i.e. that $\exists \boldsymbol{q'} \in \mathcal{O}$ such that:

$$\boldsymbol{p}^s * \boldsymbol{q}^{'} = \boldsymbol{p}^s * \boldsymbol{q}$$

Note that $\boldsymbol{\bar{p}}^{NT} > \boldsymbol{0}$ implies that we can choose $\boldsymbol{q''}$ within $\epsilon(\boldsymbol{q'})$ of $\boldsymbol{q'}$ by slightly reducing a component of $\boldsymbol{q'}$ for which the corresponding perceived price is positive. Thus, $\boldsymbol{q''} \in \mathcal{O}$, yet $\boldsymbol{p}^s * \boldsymbol{q''} < \boldsymbol{p}^s * \boldsymbol{q}$. This yields our desired contradiction. Therefore:

$$\boldsymbol{p}^s * \boldsymbol{q'} > \boldsymbol{p}^s * \boldsymbol{q} \; \forall \boldsymbol{q'} \in \mathcal{O}$$

We conclude by defining $W^s \equiv \boldsymbol{p}^s * \boldsymbol{q}$ and noting that $\forall \boldsymbol{q'} \in X$:

$$\boldsymbol{q'} \succ \boldsymbol{q} \Rightarrow \boldsymbol{q'} \in \mathcal{O} \Rightarrow \boldsymbol{p^s} > W^s$$

Therefore, the model has rationalized consumption because no preferred consumption bundle is perceived to be affordable. □

Now that we've gone through the proof, we can make a couple of observations. One, the assumption of CPC preferences is satisfied when preferences are represented by a lower semi-continuous and quasi-concave function u on $\mathbb{R}^N$, so that:

$$\forall x, y \in X : x \succeq y \Leftrightarrow u(x) \geq u(y)$$

This makes it clear that we have, in fact, generalized proposition 2. Also, note that it may be easier in practice to check to see that preferences have such a utility representation than to check that they satisfy continuous pseudo-convexity.

Two, it may appear strange that we needed to assume that X^{NT} is concave specifically if it has dimension greater than one. This is because a discrete grid for consumption of non-taxed goods can create a lumpy evaluation of non-tax expenditure, thwarting the existence of a separating hyperplane. For example, consider a consumption set $\mathbb{R}_+ \times \{0,1\}^2$, where there is one taxed good chosen continuously and two non-taxed goods chosen from $\{0,1\}$. The sticker price vector is $\bar{p} = (1,1,1)$. The consumer have preferences rationalized by the function:

$$u(\boldsymbol{q}) = q_1 + \min\{q_2, q_3\}$$

In words, the taxed good is perfect substitutes with the minimum consumption of the two non-taxed goods, which are perfect complements with each other. Consider the consumption bundle:

$$\boldsymbol{q} = (0,1,0)$$

If the agent perceived income $W^s \geq 2$, they could do better by consuming $(0,1,1)$. Supposing otherwise, if the agent perceives a positive tax-inclusive price of the taxed good, then optimally $q_1 > 0$ and $q_2 = q_3 = 0$. Finally, there is no optimal consumption bundle if $p_1^s \leq 0$. Thus, the consumption bundle cannot be rationalized.

Next, we derive our expression for the change in consumer surplus due to the tax:

Proposition 3. *Let $e(p)$ and $h(p)$ denote the expenditure function and compensated demand for the taxed good respectively at price p for the taxed good and price p^{NT} for the other good, so that the agent is minimally compensated so as to achieve utility of at least $u(\boldsymbol{q}(\bar{p}, p^{NT}, 0, W))$; formally, $e(p) = \min\{W' | u(d(p, p^{NT}, W'), d^{NT}(p, p^{NT}, W)) \geq u(\boldsymbol{q}(\bar{p}, p^{NT}, 0, W))\}$, which is well-defined by continuity of u and connectedness of the choice set. Then compensating variation due to the tax satisfies:*

$$\Delta CS = (\bar{p} + \tau - p^s)h(p^s) + e(p^s) - e(\bar{p}) \tag{10}$$

Proof: Letting W^s denote conjectured wealth when facing tax τ, local non-satiation of preferences implies that:

$$(p^s, \bar{p}^{NT}) * \boldsymbol{q}(\bar{p}, p^{NT}, \tau, W + \Delta CS) = W^s = e(p^s)$$

In words, total perceived expenditures equal perceived wealth, which must be exactly the wealth the agent would need under perceived prices to achieve the utility from before the tax. Plugging in and using the fact that $h(p^s) = q(\bar{p}, p^{NT}, \tau, W + \Delta CS)$ yields:

$$\begin{aligned}
(\bar{p} + \tau - p^s)h(p^s) &= [(\bar{p} + \tau, \bar{p}^{NT}) - (p^s, \bar{p}^{NT})] * \boldsymbol{q}(\bar{\boldsymbol{p}}, \tau, W + \Delta CS) \\
(\bar{p} + \tau - p^s)h(p^s) &= W + \Delta CS - e(p^s)
\end{aligned}$$

Rearranging and again using local non-satiation yields:

$$\Delta CS = e(p^s) - W + (\bar{p} + \tau - p^s)h(p^s) = e(p^s) - e(\bar{p}) + (\bar{p} + \tau - p^s)h(p^s)$$

□

The following lemma establishes the Compensated Law of Demand (CLD) in our setting:

Lemma 2. *For any agent i with type θ_i and any two prices p and p':*

$$p < p' \Rightarrow q(p'; \theta_i, h) \leq q(p; \theta_i, l)$$

Proof: Note that there must be values q^{NT} and $q^{NT'}$ such that:

$$(q(p; \theta_i, l), q^{NT}) \sim_{\theta_i} (q(p'; \theta_i, h), q^{NT'})$$

From local non-satiation:

$$p * q(p; \theta_i, l) + p^{NT} * q^{NT} \leq p * q(p'; \theta_i, h) + p^{NT} * q^{NT'}$$

$$p' * q(p'; \theta_i, h) + p^{NT} * q^{NT'} \leq p' * q(p; \theta_i, l) + p^{NT} * q^{NT}$$

Rearranging yields:

$$p * [q(p; \theta_i, l) - q(p'; \theta_i, h)] \leq p^{NT} * [q^{NT'} - q^{NT}] \leq p' * [q(p; \theta_i, l) - q(p'; \theta_i, h)]$$

Thus, $p' > p \Rightarrow q(p; \theta_i, l) \geq q(p'; \theta_i, h)$. □

Proposition 4. *Assume a continuously differentiable and strictly increasing aggregate supply function Q^{supply}, as well as continuously differentiable compensated demand functions h_i and subjective price functions p_i^s $\forall i$. Subjective price functions change one-for-one with sticker prices, so that:*

$$p_i^s(\bar{p}, \tau) = \bar{p} + p_i^s(0, \tau) \;\forall \bar{p} \;\forall \tau \;\forall i$$

Subjective prices also agree with sticker prices when there is no tax:

$$p_i^s(\bar{p}, 0) = \bar{p} \quad \forall \bar{p} \;\forall i$$

We implicitly define the pre-tax sticker price $\bar{p}^{old}$ by:[2]

$$Q^{supply}(\bar{p}^{old}) = \sum_i h_i(\bar{p}^{old}, \nu_i)$$

[2] $\nu_i \equiv u_i(\boldsymbol{d}_i(\boldsymbol{p}, W_i)) \;\forall i$

and the new sticker price $\bar{p}^{new}$ after the imposition of the tax τ when agents are compensated by:

$$Q^{supply}(\bar{p}^{new}) = \sum_i h_i((p_i^s(\bar{p}^{new}, \tau)), \nu_i)$$

Defining deadweight loss by:[3]

$$DWL \equiv \sum_i \Delta CS_i + \int_{\bar{p}^{new}}^{\bar{p}^{old}} Q^{supply}(p)dp - \tau \sum_i q_i^c$$

where

$$\Delta CS_i = (\bar{p}^{new} + \tau - p_i^s(\bar{p}^{new}, \tau))q_i^c + \int_{\bar{p}^{old}}^{p_i^s(\bar{p}^{new},\tau)} h_i(p, \nu_i)dp \ \forall i$$

$$q_i^c \equiv h_i(p_i^s(\bar{p}^{new}, \tau), \nu_i) \ \forall i$$

then aggregate deadweight loss has second order approximation around $\tau = 0$:

$$DWL \approx -\frac{1}{2}[\sum_i m_i \frac{\partial h_i}{\partial p} - \frac{(\sum_i m_i \frac{\partial h_i}{\partial p})^2}{\sum_i \frac{\partial h_i}{\partial p} - \frac{\partial Q^{supply}}{\partial p}}]\tau^2$$

Proof:

$$DWL = \sum_i \int_{\bar{p}^{old}}^{p_i^s(\bar{p}^{new},\tau)} h_i(p, \nu_i)dp + \int_{\bar{p}^{new}}^{\bar{p}^{old}} Q^{supply}(p)dp + \sum_i (\bar{p}^{new} - p_i^s(\bar{p}^{new}, \tau))q_i^c$$

Note that $\bar{p}^{new}$ is a function of τ. One can easily confirm that $\bar{p}^{new}|_{\tau=0} = \bar{p}^{old}$, so that deadweight loss is zero when $\tau = 0$. We can find $\frac{\partial \bar{p}^{new}}{\partial \tau}$ from the Inverse Function Theorem:[4]

$$\frac{\partial Q^{supply}}{\partial p}\frac{\partial \bar{p}^{new}}{\partial \tau} = \sum_i \frac{\partial h_i}{\partial p}[\frac{\partial p_i^s}{\partial \bar{p}^{new}}\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}] = \sum_i \frac{\partial h_i}{\partial p}[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]$$

$$\frac{\partial \bar{p}^{new}}{\partial \tau} = \frac{\sum_i \frac{\partial h_i}{\partial p}\frac{\partial p_i^s}{\partial \tau}}{\frac{\partial Q^{supply}}{\partial p} - \sum_i \frac{\partial h_i}{\partial p}}$$

We can then take the first derivative of deadweight loss with respect to the tax:

$$\begin{aligned}
\frac{\partial DWL}{\partial \tau} &= \sum_i [\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]h_i - \frac{\partial \bar{p}^{new}}{\partial \tau}Q^{supply}(\bar{p}^{new}) \\
&- \sum_i \left[\frac{\partial p_i^s}{\partial \tau}h_i + (p_i^s(\bar{p}^{new}, \tau) - \bar{p}^{new})\frac{\partial h_i}{\partial p}[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]\right] \\
&= \frac{\partial \bar{p}^{new}}{\partial \tau}\sum_i h_i - \frac{\partial \bar{p}^{new}}{\partial \tau}Q^{supply}(\bar{p}^{new}) - \sum_i \left[(p_i^s(\bar{p}^{new}, \tau) - \bar{p}^{new})\frac{\partial h_i}{\partial p}[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]\right] \\
&= -\sum_i \left[(p_i^s(\bar{p}^{new}, \tau) - \bar{p}^{new})\frac{\partial h_i}{\partial p}[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]\right]
\end{aligned}$$

Since $p_i^s(\bar{p}^{new}, 0) = \bar{p}^{new}$, it follows that

$$\left.\frac{\partial DWL}{\partial \tau}\right|_{\tau=0} = 0$$

[3]Note that $\bar{p}^{new} \leq \bar{p}^{old} \ \forall \tau \geq 0$ from the Compensated Law of Demand and the fact that supply is strictly increasing in price.

[4]This claim also uses the fact that aggregate supply is strictly increasing while aggregate compensated demand is weakly decreasing, so that there is always a unique value for $\bar{p}^{new}$.

Obtaining the second derivative would be straightforward if $h_i \in \mathbb{C}^2\ \forall i$. Instead, we find the second derivative at $\tau = 0$ from the definition:

$$\frac{\partial^2 DWL}{\partial \tau^2}\bigg|_{\tau=0} = \lim_{\tau\to 0} -\frac{\sum_i \left[(p_i^s(\bar{p}^{new},\tau) - \bar{p}^{new})\frac{\partial h_i}{\partial p}[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}]\right]}{\tau}$$

Note that continuity of $\frac{\partial p_i^s}{\partial \tau}$ with respect to τ for all agents implies that $\frac{\partial \bar{p}^{new}}{\partial \tau}$ is continuous. Since $\frac{\partial Q^{supply}}{\partial p}$ and $\frac{\partial h_i}{\partial p}$ $\forall i$ are also continuous:

$$\begin{aligned}
\frac{\partial^2 DWL}{\partial \tau^2}\bigg|_{\tau=0} &= -\sum_i \frac{\partial h_i}{\partial p}\bigg|_{\tau=0}[\frac{\partial \bar{p}^{new}}{\partial \tau}\bigg|_{\tau=0} + \frac{\partial p_i^s}{\partial \tau}\bigg|_{\tau=0}]\lim_{\tau\to 0}\frac{(p_i^s(\bar{p}^{new},\tau) - \bar{p}^{new})}{\tau} \\
&= -\sum_i \frac{\partial h_i}{\partial p}\bigg|_{\tau=0}[\frac{\partial \bar{p}^{new}}{\partial \tau}\bigg|_{\tau=0} + \frac{\partial p_i^s}{\partial \tau}\bigg|_{\tau=0}]\frac{\partial p_i^s}{\partial \tau}\bigg|_{\tau=0}
\end{aligned}$$

Using the fact that $m_i \equiv \frac{\partial p_i^s}{\partial \tau}|_{\tau=0}$, we can note that:

$$\frac{\partial \bar{p}^{new}}{\partial \tau}|_{\tau=0} = \frac{\sum_i m_i \frac{\partial h_i}{\partial p}}{\frac{\partial Q^{supply}}{\partial p} - \sum_i \frac{\partial h_i}{\partial p}}$$

and so:

$$\frac{\partial^2 DWL}{\partial \tau^2}|_{\tau=0} = -[\sum_i m_i^2 \frac{\partial h_i}{\partial p} + \frac{(\sum_i m_i \frac{\partial h_i}{\partial p})^2}{\frac{\partial Q^{supply}}{\partial p} - \sum_i \frac{\partial h_i}{\partial p}}]$$

Now we can find the second order approximation for deadweight loss:

$$DWL \approx DWL|_{\tau=0} + \frac{\partial DWL}{\partial \tau}|_{\tau=0}\tau + \frac{1}{2}\frac{\partial^2 DWL}{\partial \tau^2}|_{\tau=0}\tau^2$$

$$DWL \approx -\frac{1}{2}[\sum_i m_i^2 \frac{\partial h_i}{\partial p} + \frac{(\sum_i m_i \frac{\partial h_i}{\partial p})^2}{\frac{\partial Q^{supply}}{\partial p} - \sum_i \frac{\partial h_i}{\partial p}}]\tau^2$$

□

A.2 Additional results and proofs from section 3

The upper and lower bounds use the following lemma:

Lemma 3. *For any agent i with type θ_i and any two pairs (p, ζ_i) and (p', ζ_i'):*

$$dwl(p';\theta_i,\zeta_i) \geq dwl(p;\theta_i,\zeta_i') - (p - \bar{p})(q(p';\theta_i,\zeta_i') - q(p;\theta_i,\zeta_i)).$$

Proof: Note from the definition of the expenditure function and optimal compensated consumption vectors $\boldsymbol{q}$ and $\boldsymbol{q'}$ for price vectors (p, p^{NT}) and (p', p^{NT}) respectively:

$$e(p') - e(p) = (p', p^{NT}) * \boldsymbol{q'} - (p, p^{NT}) * \boldsymbol{q} \geq (p', p^{NT}) * \boldsymbol{q'} - (p, p^{NT}) * \boldsymbol{q'} = (p' - p)q(p';\theta_i,\zeta_i')$$

Plugging in yields:

$$\begin{aligned}
dwl(p';\theta_i) &= [e(p') - e(\bar{p})] - (p' - \bar{p})q(p';\theta_i,\zeta_i') \\
&= [e(p') - e(p)] + [e(p) - e(\bar{p})] - [(p' - p) + (p - \bar{p})]q(p';\theta_i,\zeta_i') \\
&\geq (p' - p)q(p';\theta_i,\zeta_i') + e(p) - e(\bar{p}) - [(p' - p) + (p - \bar{p})]q(p';\theta_i,\zeta_i') \\
&= e(p) - e(\bar{p}) - (p - \bar{p})q(p';\theta_i,\zeta_i') \\
&= dwl(p) + (p - \bar{p})q(p;\theta_i,\zeta_i) - (p - \bar{p})q(p';\theta_i,\zeta_i') \\
&= dwl(p;\theta_i) - (p - \bar{p})(q(p';\theta_i,\zeta_i') - q(p;\theta_i,\zeta_i))
\end{aligned}$$

See also appendix figure 5 for a graphical demonstration. □

Proof of Proposition 1: From lemma 2 and prices being bound away from zero, we can always find a value of $\hat{p}^s$ such that:

$$\int_{\theta_i} q(\hat{p}^s;\theta_i,l)dF_\theta(\theta_i) \le \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \le \int_{\theta_i} q(\hat{p}^s;\theta_i,h)dF_\theta(\theta_i)$$

Pick $\lambda \in [0,1]$ such that:

$$\lambda \int\limits_{\theta_i} q(\hat{p}^s;\theta_i,h)dF_\theta(\theta_i) + (1-\lambda)\int\limits_{\theta_i} q(\hat{p}^s;\theta_i,l)dF_\theta(\theta_i) = \int\limits_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

Define $F'_{\theta,\zeta}$ such that $F'_\theta = F_\theta$ and $\zeta = h$ with probability λ, $\zeta = l$ with probability $1-\lambda$, $\theta \perp \zeta$. Then:

$$\begin{aligned}\int_{\theta_i,\zeta_i} q(\hat{p}^s;\theta_i,\zeta_i)dF_\theta(\theta_i) &= \int_{\theta_i}[\lambda q(\hat{p}^s;\theta_i,h)+(1-\lambda)q(\hat{p}^s;\theta_i,l)]dF_\theta(\theta_i)\\ &= \int_{\theta_i,\zeta_i} q(pl_i^s;\theta_i,\zeta_i)dF'_{p^s,\theta,\zeta}(\theta_i,\zeta_i)\end{aligned}$$

□

Proof of theorem 1: From lemma 3:

$$\begin{aligned}&\int_{p_i^s,\theta_i,\zeta_i}[dwl(p_i^s;\theta_i,\zeta_i)+(\hat{p}^s-\bar{p})q(p_i^s;\theta_i,\zeta_i)]dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &\qquad \ge \int_{p_i^s,\theta_i,\zeta_i}[dwl(\hat{p}^s;\theta_i,\zeta_i)+(\hat{p}^s-\bar{p})q(\hat{p}^s;\theta_i,\zeta_i)]dF'_{\theta,\zeta}(\theta_i,\zeta_i)\end{aligned}$$

But note from the rationalizability of the data that:

$$(\hat{p}^s-\bar{p})\int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) = (\hat{p}^s-\bar{p})\int_{\theta_i,\zeta_i} q(\hat{p}^s;\theta_i,\zeta_i)dF'_{\theta,\zeta}(\theta_i,\zeta_i)$$

We can thus conclude that:

$$\int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \ge \int_{\theta_i,\zeta_i} dwl(\hat{p}^s;\theta_i,\zeta_i)dF_{\theta,\zeta}(\theta_i,\zeta_i)$$

□

Proof of theorem 2: From lemma 3 and rationalizability of the data:

$$\begin{aligned}&\int\limits_{p_i^s,\theta_i}[dwl(p^b(p_i^s);\theta_i)+(p_i^s-\bar{p})q(p^b(p_i^s);\theta_i,\zeta_i)]dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &\qquad \ge \int\limits_{p_i^s,\theta_i,\zeta_i}[dwl(p_i^s;\theta_i)+(p_i^s-\bar{p})q(p_i^s;\theta_i,\zeta_i)]dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\end{aligned}$$

$$\begin{aligned}&\int\limits_{p_i^s,\theta_i}[dwl(p^b(p_i^s);\theta_i)+p_i^s q(p^b(p_i^s);\theta_i,\zeta_i)]dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &\qquad \ge \int\limits_{p_i^s,\theta_i,\zeta_i}[dwl(p_i^s;\theta_i)+p_i^s q(p_i^s;\theta_i,\zeta_i)]dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\end{aligned}$$

$$\begin{aligned}&\int\limits_{p_i^s,\theta_i}[dwl(p^b(p_i^s);\theta_i,\zeta_i)+(p_i^s-\tilde{p}^s)q(p^b(p_i^s);\theta_i,\zeta_i)]dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\ &\qquad \ge \int\limits_{p_i^s,\theta_i,\zeta_i}[dwl(p_i^s;\theta_i)+(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)]dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\end{aligned}$$

Rearranging yields:

$$\int_{p_i^s,\theta_i,\zeta_i} dwl(p^b(p_i^s);\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \geq \int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$
$$-\int_{p_i^s,\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p^b(p_i^s);\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) + \int_{p_i^s,\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

We can show from lemma 2 and $p_i^s \in [\bar{p}, \bar{p}+\bar{m}\tau] = \mathcal{P}\ \forall i$ that the term on the second line is non-negative. Formally for any $p_i^s \in (\bar{p}, \bar{p}+\bar{m}\tau), \theta_i, \zeta_i, \zeta_i'$:

$$p_i^s > \tilde{p}^s \Rightarrow p^b(p_i^s) > \tilde{p}^s \Rightarrow q(p^b(p_i^s);\theta_i,\zeta_i') \leq q(p_i^s;\theta_i,\zeta_i)$$
$$p_i^s \leq \tilde{p}^s \Rightarrow p^b(p_i^s) < \tilde{p}^s \Rightarrow q(p^b(p_i^s);\theta_i,\zeta_i') \geq q(p_i^s;\theta_i,\zeta_i)$$

Either way:

$$(p_i^s-\tilde{p}^s)q(p^b(p_i^s);\theta_i,\zeta_i') \leq (p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)$$

Thus:

$$\begin{aligned}
\int_{p_i^s,\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) &= \int_{p_i^s\in int(\mathcal{P}),\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\
&+\int_{p_i^s\in\{\bar{p},\bar{p}+\bar{m}\tau\},\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\
&\geq \int_{p_i^s\in int(\mathcal{P}),\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p^b(p_i^s);\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\
&+\int\limits_{p_i^s\in\{\bar{p},\bar{p}+\bar{m}\tau\},\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)\\
&= \int_{p_i^s,\theta_i,\zeta_i}(p_i^s-\tilde{p}^s)q(p_i^s;\theta_i,\zeta_i)dF''_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)
\end{aligned}$$

$$\int\limits_{p_i^s,\theta_i,\zeta_i} dwl(p^b(p_i^s);\theta_i,\zeta_i)dF_{p^s,\theta}(p_i^s,\theta_i) \geq \int\limits_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta}(p_i^s,\theta_i,\zeta_i)$$

□

Before stating and proving theorem 3, we note that deadweight loss is bounded by the product of the reduction in demand and $\bar{m}\tau$.

Lemma 4. *If $p_i^s \in [\bar{p}, \bar{p}+\bar{m}\tau]$, then $dwl(p_i^s;\theta_i,\zeta_i) \leq [q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)]\bar{m}\tau\ \forall\theta_i,\zeta_i$.*

Proof: Using lemma 3:

$$0 = dwl(\bar{p};\theta_i,\zeta_i) \geq dwl(p_i^s;\theta_i,\zeta_i) - (p_i^s-\bar{p})[q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)]$$
$$dwl(p_i^s;\theta_i,\zeta_i) \leq [q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)](p_i^s-\bar{p}) \leq [q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)]\bar{m}\tau.$$

□

Next, we state and prove theorem 3. It says that the maximal value of deadweight loss consistent with the data and knowledge of F_θ^* is given by having some agents perceive the highest possible price and some others perceive the lowest possible price. It achieves this by assigning the good there where it will generate the most deadweight loss, while holding aggregate demand constant. The resulting demand function, $\tilde{q}_{\Delta,\gamma}(\theta_i)$, is such that those for whom the ratio of deadweight loss[5] to change in quantity exceeds Δ perceive the highest price, those with such a ratio below Δ perceive the lowest possible price, and those with ratio equal to Δ are split between perceiving the high and low price in a way that rationalizes demand. Those who perceive the high (low) price consume the least (most) possible consistent with their perceptions.

[5]Note that when $p^s = \bar{p}$, the tax is effectively lump-sum and so there is no deadweight loss.

Theorem 3. *There exist values $\Delta \in [0, \bar{m}\tau]$ and $\gamma \in [0,1]$ such that:*

$$\int_{p_i^s,\theta_i} \tilde{q}_{\Delta,\gamma}(\theta_i) dF^*_{p^s,\theta_i} = \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i) dF^*_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

where:

$$\tilde{q}_{\Delta,\gamma}(\theta_i) = \left[\mathbb{I}\left(\frac{dwl(\bar{p}+\bar{m}\tau;\theta_i,l)}{q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)} > \Delta\right) + \gamma\mathbb{I}\left(\frac{dwl(\bar{p}+\bar{m}\tau;\theta_i,l)}{q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)} = \Delta\right)\right] q(\bar{p}+\bar{m}\tau;\theta_i,l)$$
$$+ \left[\mathbb{I}\left(\frac{dwl(\bar{p}+\bar{m}\tau;\theta_i,l)}{q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)} < \Delta\right) + (1-\gamma)\mathbb{I}\left(\frac{dwl(\bar{p}+\bar{m}\tau;\theta_i,l)}{q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)} = \Delta\right)\right] q(\bar{p};\theta_i,h)$$

Of course, if $q(\bar{p};\theta_i,h) = q(\bar{p}+\bar{m}\tau;\theta_i,l)$, then $\tilde{q}_{\Delta,\gamma}(\theta_i) = q(\bar{p};\theta_i,h)$. Furthermore, under assumption 1, for any $F_{p^s,\theta,\zeta}$ that rationalizes the data such that $F_\theta = F^*_\theta$:

$$\int_{p_i^s,\theta_i} \frac{\tilde{q}_{\Delta,\gamma}(\theta_i) - q(\bar{p};\theta_i,h)}{q(\bar{p}+\bar{m}\tau;\theta_i,l) - q(\bar{p};\theta_i,h)} dwl(\bar{p}+\bar{m}\tau;\theta_i,l) dF^*_{p^s,\theta}(p_i^s,\theta_i) \geq \int_{p_i^s,\theta_i,\zeta_i} dwl(p_i^s;\theta_i,\zeta_i) dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

where the integrand on the left-hand side is defined as zero for any θ_i such that $q(\bar{p}+\bar{m}\tau;\theta_i,l) = q(\bar{p};\theta_i,h)$.

The intuition is for the Δ term is straightforward. The econometrician observes the reduction in aggregate demand due to the tax. In searching for the explanation of that reduction in demand that maximizes deadweight loss, one should assign the reduction in quantity demanded to those for whom that allocation yields the greatest deadweight loss. Following this procedure, there is a cutoff value Δ which describes the amount of deadweight loss obtained relative to the reduction in quantity demanded sufficient to warrant the assignment of subjective tax-inclusive price $p_i^s = \bar{p} + \bar{m}\tau$ to that agent.

The idea behind the tie-breaking provision is that those individuals who perceive the high price should reduce their consumption as much as possible to maximize deadweight loss; those who perceive the sticker price should maximize their consumption to permit even more individuals to perceive the high price.

Proof of Theorem 3: The outline of the proof is as follows. First, we use lemma 4 to show that the maximal deadweight loss consistent with aggregate demand and F^*_θ comes from a data-generating process in which agents perceiving the price $\bar{p} + \bar{m}\tau$ choose the lowest quantity consistent with preference maximization, whereas the other agents choose the largest such quantity. Then, we show that distributions satisfying such a property yield deadweight loss no larger than the proposed distribution, which exists.

First, consider an arbitrary distribution $F_{p^s,\theta,\zeta}$ (yielding well-defined aggregate demand and deadweight loss) such that $F_\theta = F^*_\theta$ and:

$$F_{p^s} = \begin{cases} 0 & p_i^s < \bar{p} \\ F_{p^s}(\bar{p}) & p_i^s \in [\bar{p}, \bar{p}+\bar{m}\tau) \\ 1 & p_i^s \geq \bar{p}+\bar{m}\tau \end{cases}$$

In words, the above expression says that the support of p^s is contained in $\{\bar{p}, \bar{p}+\bar{m}\tau\}$. By theorem 2, the maximal value of deadweight loss consistent with aggregate demand and F^*_θ must satisfy this property. Consider some value $\rho \in [0,1]$ such that:

$$\begin{aligned} &\rho \int_{\theta_i} q(\bar{p}+\bar{m}\tau;\theta_i,l) dF_{\theta|p^s\neq\bar{p}}(\theta_i) + [1-\rho] \int_{\theta_i} q(\bar{p};\theta_i,h) dF_{\theta|p^s=\bar{p}}(\theta_i) \\ &= \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i) dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \end{aligned} \quad (11)$$

Such a value of ρ must exist by the Intermediate Value Theorem, since by the definition of l and h and the CLD as expressed in lemma 2:

$$\int_{\theta_i} q(\bar{p}+\bar{m}\tau;\theta_i,l) dF_{\theta|p^s\neq\bar{p}}(\theta_i) \leq \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i) dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \quad \leq \int_{\theta_i} q(\bar{p};\theta_i,h) dF_{\theta|p^s=\bar{p}}(\theta_i)$$

In words, we are constructing an alternative distribution that rationalizes aggregate demand such that $p^s = \bar{p} + \bar{m}\tau$ and $\zeta = l$ with probability ρ, and otherwise $p^s = \bar{p}$ and $\zeta = h$. We now show that this alternate distribution yields at least as much deadweight loss, thus showing that the maximal value of deadweight loss consistent with aggregate demand and F^*_θ must arise from a distribution in which almost surely $(p^s, \zeta) = (\bar{p}, h)$ or $(p^s, \zeta) = (\bar{p} + \bar{m}\tau, l)$.

From the definition of deadweight loss:

$$\int_{\theta_i,\zeta_i} \bar{m}\tau[q(\bar{p};\theta_i,\zeta_i) - q(\bar{p};\theta_i,l)]dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$
$$= \int_{\theta_i,\zeta_i} [dwl(\bar{p}+\bar{m}\tau;\theta_i,l) - dwl(\bar{p};\theta_i,\zeta_i)]dF_{\theta,\zeta|p^s\neq\bar{p}}$$

From here, the definition of l, and using the fact that $dwl(\bar{p};\theta_i,\zeta_i) = 0\ \forall\theta_i,\zeta_i$, we have that $\rho \geq 1 - F_{p^s}(\bar{p})$ implies that:

$$\rho\int_{\theta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,l)dF_{\theta|p^s\neq\bar{p}}(\theta_i) + (1-\rho)\int_{\theta_i} dwl(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i)$$
$$= \rho\int_{\theta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,l)dF_{\theta|p^s\neq\bar{p}}(\theta_i)$$
$$\geq [1-F_{p^s}(\bar{p})]\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$
$$= [1-F_{p^s}(\bar{p})]\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} dwl(\bar{p};\theta_i,\zeta_i)dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i)$$
$$= \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta}(\theta_i,\zeta_i)$$

Where the inequality follows from the fact that $\rho \geq 1 - F_{p^s}(\bar{p})$ by assumption, and the fact that $dwl(\bar{p} + \bar{m}\tau;\theta_i,\zeta_i)$ and the definition of l. This shows that whenever $\rho \geq 1 - F_{p^s}(\bar{p})$, the proposed alternative distribution yields at least as much deadweight loss. Now suppose instead $\rho < 1 - F_{p^s}(\bar{p})$. From lemma 4:

$$\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$
$$\geq \bar{m}\tau\int_{\theta_i,\zeta_i} [q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)]\bar{m}\tau dF_{\theta_i,\zeta_i|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$

In addition, we find it convenient to rewrite the aggregate demand-rationalizing equation as:

$$\int_{p^s_i,\theta_i,\zeta_i} q(p^s_i;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p^s_i,\theta_i,\zeta_i)$$
$$= [1-F_{p^s}(\bar{p})]\int_{\theta_i,\zeta_i} q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} q(\bar{p};\theta_i,\zeta_i)dF_{\theta|p^s=\bar{p}}(\theta_i,\zeta_i)$$

And so, using equation 11 and rearranging terms,

$$\rho\int_{\theta_i,\zeta_i} [q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]dF_{\theta,\zeta}(\theta_i,\zeta_i)$$
$$= (1-\rho)\int_{\theta_i} q(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i) - [1-F_{p^s}(\bar{p})-\rho]\int_{\theta_i,\zeta_i} q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$
$$- F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i)$$

Thus, plugging in and using lemma 4:

$$
\begin{aligned}
&\rho \int_{\theta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,l)dF_{\theta|p^s\neq\bar{p}}(\theta_i) + (1-\rho)\int_{\theta_i} dwl(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i) \\
&= \rho \int_{\theta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,l)dF_{\theta|p^s\neq\bar{p}}(\theta_i) \\
&= \rho \int_{\theta_i,\zeta_i} \big[dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i) - q(\bar{p}+\bar{m}\tau;\theta_i,l) + q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)\big]dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&= \rho \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&+ \rho \int_{\theta_i,\zeta_i} [q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]dF_{\theta,\zeta}(\theta_i,\zeta_i) \\
&= \rho \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + (1-\rho)\int_{\theta_i} q(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i) \\
&- [1-F_{p^s}(\bar{p})-\rho]\int_{\theta_i,\zeta_i} q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&- F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} q(\bar{p};\theta_i,\zeta_i)dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i) \\
&= \rho \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + F_{p^s}(\bar{p})\int_{\theta_i} q(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i) \\
&+ [1-F_{p^s}(\bar{p})-\rho]\int_{\theta_i,\zeta_i} \big[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)\big]dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&- F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} q(\bar{p};\theta_i,\zeta_i)dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i) \\
&\geq \rho \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + F_{p^s}(\bar{p})\int_{\theta_i} q(\bar{p};\theta_i,h)dF_{\theta|p^s=\bar{p}}(\theta_i) \\
&+ [1-F_{p^s}(\bar{p})-\rho]\bar{m}\tau\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&- F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} q(\bar{p};\theta_i,\zeta_i)dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i) \\
&\geq \rho \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&+ [1-F_{p^s}(\bar{p})-\rho]\bar{m}\tau\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) \\
&= [1-F_{p^s}(\bar{p})]\int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i)dF_{\theta,\zeta|p^s\neq\bar{p}}(\theta_i,\zeta_i) + F_{p^s}(\bar{p})\int_{\theta_i,\zeta_i} dwl(\bar{p};\theta_i,\zeta_i)dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i)
\end{aligned}
$$

Thus, we know that the maximal deadweight loss consistent with aggregate demand and F^*_θ is generated by a distribution in which with probability one either $(p^s,\zeta) = (\bar{p},h)$ or $(p^s,\zeta) = (\bar{p}+\bar{m}\tau,l)$. We refer to distributions of this sort as *binary* distributions.

Now, we show that the proposed distribution maximizes deadweight loss among all binary distributions, and thus among all distributions, that rationalize aggregate demand such that $F_\theta = F^*_\theta$. Towards that end, we first show that the proposed distribution exists. Note by lemma 4 and the CLD as in lemma 2:

$$
\int_{\theta_i} \tilde{q}_{\bar{m}\tau,1}(\theta_i)dF^*_{p^s,\theta}(p^s_i,\theta_i) \leq \int_{p^s_i,\theta_i,\zeta_i} q(p^s_i;\theta_i,\zeta_i)dF^*_{p^s,\theta,\zeta}(p^s_i,\theta_i,\zeta_i) \leq \int_{\theta_i} \tilde{q}_{0,0}(\theta_i)dF^*_{p^s,\theta}(p^s_i,\theta_i)
$$

In words, aggregate demand is contained between when all agents perceive a high price and have type h and when all agents perceive a low price and have type l. Furthermore, one can confirm that for any $\Delta,\Delta',\gamma,\gamma'$

such that $0 \leq \Delta < \Delta' \leq \bar{m}\tau$ and $0 \leq \gamma < \gamma' \leq 1$:

$$\int_{\theta_i} \tilde{q}_{\Delta,\gamma'}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i) \leq \int_{\theta_i} \tilde{q}_{\Delta,\gamma}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i)$$

$$\int_{\theta_i} \tilde{q}_{\Delta',\gamma'}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i) \geq \int_{\theta_i} \tilde{q}_{\Delta,\gamma}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i)$$

Thus, we can pick Δ such that:

$$\int\limits_{\theta_i} \tilde{q}_{\Delta,1}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i) \leq \int\limits_{p^s_i,\theta_i,\zeta_i} q(p^s_i; \theta_i, \zeta_i) dF^*_{p^s,\theta,\zeta}(p^s_i, \theta_i, \zeta_i) \leq \int\limits_{\theta_i} \tilde{q}_{\Delta,0}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i)$$

If both sides hold with equality, we can define γ arbitrarily. Otherwise, we define γ so that the market clears:

$$\gamma \equiv \frac{\int_{p^s_i,\theta_i,\zeta_i} q(p^s_i; \theta_i, \zeta_i) dF^*_{p^s,\theta,\zeta}(p^s_i, \theta_i, \zeta_i) - \int_{\theta_i} \tilde{q}_{\Delta,0}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i)}{\int_{\theta_i} \tilde{q}_{\Delta,1}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i) - \int_{\theta_i} \tilde{q}_{\Delta,0}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i)}$$

We now have the values Δ and γ such that the market clears. Suppressing Δ and γ subscripts from $\tilde{q}$, we can say that:

$$\int_{\theta_i} \tilde{q}(\theta_i) dF^*_{p^s,\theta}(p^s_i, \theta_i) = \int_{p^s_i,\theta_i,\zeta_i} q(p^s_i; \theta_i, \zeta_i) dF^*_{p^s,\theta,\zeta}(p^s_i, \theta_i, \zeta_i)$$

Finally, to show that the proposed distribution maximizes deadweight loss, consider arbitrary binary distribution $F_{p^s,\theta,\zeta}$ that rationalizes aggregate demand. Defining $\mathbb{P}_F(p^s \neq \bar{p}|\theta_i) \equiv 1 - F_{p^s|\theta=\theta_i}(\bar{p} + \bar{m}\tau)$ as the probability that $(p^s, \zeta) = (\bar{p} + \bar{m}\tau, l)$ conditional on θ_i, rationalizing aggregate demand with $F_\theta = F^*_\theta$ means that:

$$\int_{\theta_i} \left[\mathbb{P}_F(p^s \neq \bar{p}|\theta_i)\ q(\bar{p} + \bar{m}\tau; \theta_i, l) + F_{p^s|\theta=\theta_i}(\bar{p})\ q(\bar{p}; \theta_i, h)\right] dF^*_\theta(\theta_i)$$
$$= \int_{p^s_i,\theta_i,\zeta_i} q(p^s_i; \theta_i, \zeta_i) dF^*_{p^s,\theta,\zeta}(p^s_i, \theta_i, \zeta_i)$$

We can now write the difference in generated values of aggregate deadweight loss as:

$$\begin{aligned}
&\int_{\theta_i} \Big[\frac{\tilde{q}(\theta_i) - q(\bar{p}; \theta_i, h)}{q(\bar{p} + \bar{m}\tau; \theta_i, l) - q(\bar{p}; \theta_i, h)} - \mathbb{P}_F(p^s \neq \bar{p}|\theta_i)\Big] dwl(\bar{p} + \bar{m}\tau; \theta_i, l) dF^*_\theta(\theta_i) \\
&\quad = \int_{\theta_i : dwl(\bar{p}+\bar{m}\tau) > \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [1 - \mathbb{P}_F(p^s \neq \bar{p}|\theta_i)] dwl(\bar{p} + \bar{m}\tau; \theta_i) dF^*_\theta(\theta_i) \\
&\quad + \int_{\theta_i : dwl(\bar{p}+\bar{m}\tau) = \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [\gamma - \mathbb{P}_F(p^s \neq \bar{p}|\theta_i)] dwl(\bar{p} + \bar{m}\tau; \theta_i) dF^*_\theta(\theta_i) \\
&\quad - \int_{\theta_i : dwl(\bar{p}+\bar{m}\tau) < \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} \mathbb{P}_F(p^s \neq \bar{p}|\theta_i) dwl(\bar{p} + \bar{m}\tau; \theta_i) dF^*_\theta(\theta_i) \\
&\quad \geq \Delta \int\limits_{\theta_i : dwl(\bar{p}+\bar{m}\tau) > \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [1 - \mathbb{P}_F(p^s \neq \bar{p}|\theta_i)]\ [q(\bar{p}; \theta_i, h) - q(\bar{p} + \bar{m}\tau; \theta_i, l)] dF^*_\theta(\theta_i) \\
&\quad + \Delta \int\limits_{\theta_i : dwl(\bar{p}+\bar{m}\tau) = \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [\gamma - \mathbb{P}_F(p^s \neq \bar{p}|\theta_i)]\ [q(\bar{p}; \theta_i, h) - q(\bar{p} + \bar{m}\tau; \theta_i, l)] dF^*_\theta(\theta_i) \\
&\quad - \Delta \int\limits_{\theta_i : dwl(\bar{p}+\bar{m}\tau) < \Delta[q(\bar{p};\theta_i,h) - q(\bar{p}+\bar{m}\tau;\theta_i,l)]} \mathbb{P}_F(p^s \neq \bar{p}|\theta_i) [q(\bar{p}; \theta_i, h) - q(\bar{p} + \bar{m}\tau; \theta_i, l)] dF^*_\theta(\theta_i)
\end{aligned}$$

We complete the proof by showing the right-hand side of the last inequality is zero. Since both distributions rationalize the same aggregate demand:

$$\begin{aligned}
&\int_{\theta_i: dwl(\bar{p}+\bar{m}\tau)>\Delta[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]} q(\bar{p}+\bar{m}\tau;\theta_i,l)dF^*_\theta(\theta_i) \\
&+\int_{\theta_i: dwl(\bar{p}+\bar{m}\tau)=\Delta[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [\gamma q(\bar{p}+\bar{m}\tau;\theta_i,l)+(1-\gamma)q(\bar{p};\theta_i,h)]dF^*_\theta(\theta_i) \\
&+\int_{\theta_i: dwl(\bar{p}+\bar{m}\tau)<\Delta[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]} q(\bar{p};\theta_i,h)dF^*_\theta(\theta_i) \\
&=\int_{\theta_i} \big[\mathbb{P}_F(p^s\neq p^s|\theta_i)\ [q(\bar{p}+\bar{m}\tau;\theta_i,l)-q(\bar{p};\theta_i,h)]+q(\bar{p};\theta_i,h)\big]dF^*_\theta(\theta_i)
\end{aligned}$$

Subtracting both sides from $\int_{\theta_i} q(\bar{p};\theta_i,h)dF^*_\theta(\theta_i)$ yields:

$$\begin{aligned}
&\int_{\theta_i: dwl(\bar{p}+\bar{m}\tau)>\Delta[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]} [q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]dF^*_\theta(\theta_i) \\
&+\int_{\theta_i: dwl(\bar{p}+\bar{m}\tau)=\Delta[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]} \gamma[q(\bar{p};\theta_i,h)-q(\bar{p}+\bar{m}\tau;\theta_i,l)]dF^*_\theta(\theta_i) \\
&=\int_{\theta_i} \mathbb{P}_F(p^s\neq p^s|\theta_i)\ [q(\bar{p}+\bar{m}\tau;\theta_i,l)-q(\bar{p};\theta_i,h)]dF^*_\theta(\theta_i)
\end{aligned}$$

Finally, subtracting the right-hand side from the left-hand size and multiplying by zero yields the desired result. Thus:

$$\int_{\theta_i} \big[\frac{\tilde{q}(\theta_i)-q(\bar{p};\theta_i,h)}{q(\bar{p}+\bar{m}\tau;\theta_i,l)-q(\bar{p};\theta_i,h)}-\mathbb{P}_F(p^s\neq\bar{p}|\theta_i)\big]dwl(\bar{p}+\bar{m}\tau;\theta_i,l)dF^*_\theta(\theta_i)=0$$

In words, deadweight loss from the proposed distribution is at least as great as the deadweight loss from any binary distribution that also rationalizes aggregate demand and with the true distribution of preference types. From the first part of the proof, any distribution that rationalized aggregate demand and had the support of perceived prices contained in $\partial d\mathcal{P}$ yielded deadweight loss no greater than what one could obtain with a binary distribution that rationalized aggregate demand with $F_\theta = F^*_\theta$. Theorem 2 noted that any distribution that rationalized aggregate demand with $F_\theta = F^*_\theta$ yielded deadweight loss no greater than that one could obtain with a distribution that had the support of perceived prices contained in $\partial d\mathcal{P}$, rationalized aggregate demand, and had $F_\theta = F^*_\theta$. Therefore, any distribution that rationalizes aggregate demand and with $F_\theta = F^*_\theta$ yields deadweight loss no greater than the proposed distribution. $\square$

B Details on Application of Linear Model

We use data gathered by CLK (2009) on the aggregate consumption of beer in U.S. states between 1970 and 2003, and cross-sectional data gathered by Goldin and Homonoff (2013) on tobacco consumption between 1984 and 2000. We translate their two models (in logs) to our linear specification. In particular, we are interested in estimating models like the one in equation 8.

In the case of aggregate beer consumption, we follow CLK (2009) in using a specification in first differences, and so we estimate regressions of the type:

$$\Delta y_{st} = \alpha + \beta\Delta\tau^e_{st} + \tilde{\beta}\Delta\tau^s_{st} + \gamma X_{st} + \epsilon_{st}$$

where y_{st} represents per-capita consumption of beer, in gallons, for state s at time t, τ^e represents excise taxes on beer (included in sticker price), τ^s represents sales taxes (non-salient), X is a vector of controls, and ϵ is an i.i.d. error term. All taxes are expressed in dollar amounts.

For each linear specification, we compute $\hat{m} = \frac{\tilde{\beta}}{\beta}$, which gives us the ratio of upper bound of deadweight loss to lower bound of deadweight loss (assuming that maximal attention, $\bar{m} = 1$). Results are presented in table 1. We also estimate a number of other specifications, again following CLK (2009), presented in table 2. These are meant to address concerns for spurious results – in particular, it could be the case that consumers react differently to the two tax rates because while sales taxes affect a variety of goods, excise taxes on beer affect only beer prices. The second last column of table 2 shows estimates for a regression only for those states that exempt food (a likely substitute of beer) from sales tax, demonstrating that even in this restricted sample beer consumption is quite insensitive to sales tax. Finally, the last column addresses the potential concern that people might be substituting toward other alcoholic beverages when they face a beer tax increase, and not when they face a sales tax increase. As we can see, the share of ethanol people consume in the form of beer is quite insensitive to either tax rate.

We repeat the exercise for Goldin and Homonoff (2013), who have a similar set-up with individual-level, cross-sectional data on cigarette consumption. Even though this is not aggregate data, estimating a linear model that only measures average effects effectively leaves the analysis of section 4 unchanged. We again follow the original authors of the paper when we estimate the equation:

$$c_{ist} = \alpha + \beta\tau^e_{st} + \tilde{\beta}\tau^s_{st} + \gamma X_{st} + \delta Z_{ist} + \varepsilon_{ist}$$

where now c_{ist} stands for tobacco consumption, in average cigarettes per day, for individual i from state s in period t, τ^e, τ^s, and X_{st} should be interpreted as before, and Z_{ist} is a vector of individual-level controls. All the details can be found in the original paper. Results in table 3 showcase a number of different specifications, including several sets of fixed-effects, all following Goldin and Homonoff (2013).

	Baseline	Business cycle	Alcohol regulations	Region trends
Δ(excise tax)	-0.966	-0.875	-0.808	-0.715
	(0.4)	(0.393)	(0.394)	(0.394)
Δ(sales tax)	-0.305	-0.113	-0.114	-0.241
	(0.708)	(0.698)	(0.699)	(.7)
Δ(population)	-0.0002	-.0002	-0.0001	-0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Δ(income per cap.)		0.0002	0.0001	0.0002
		(0.00006)	(0.00006)	(0.00006)
Δ(unemployment)		-.094	-0.093	-0.093
		(.026)	(0.026)	(0.026)
Alcohol reg. controls			X	X
Year FE	X	X	X	X
Region FE				X
$\hat{m}$	0.316	0.129	0.141	0.338
	(0.743)	(0.8)	(0.866)	(0.996)
Sample size	1,607	1,487	1,487	1,487

Table 1: Estimating $\hat{m}$ with several sets of controls, following the specifications in CLK (2009) in the context of a linear model. Standard errors in parentheses.

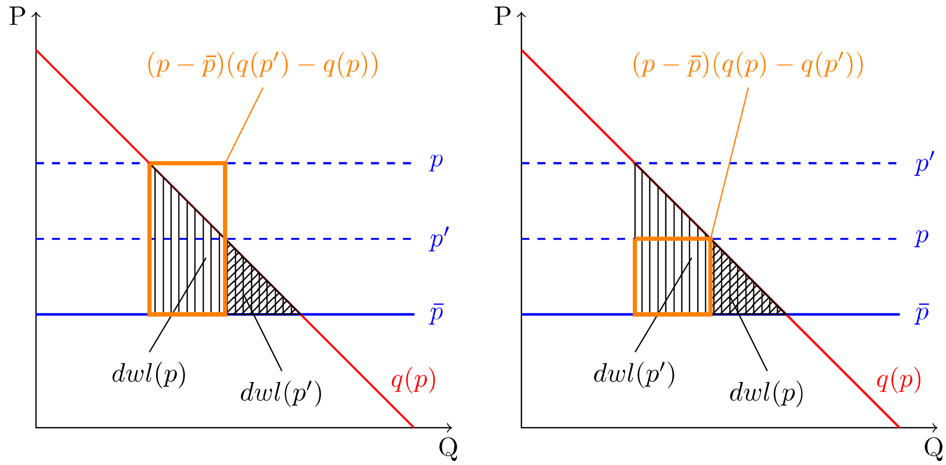
	Policy IV for excise tax	3-Year differences	Food exempt	Dep. var.: share of ethanol from beer
Δ(excise tax)	-0.808	-2.092	-1.114	0.036
	(0.395)	(0.897)	(1.174)	(0.006)
Δ(sales tax)	-0.114	-0.131	-0.449	0.018
	(0.699)	(0.826)	(0.757)	(0.011)
Δ(population)	-0.0001	-0.002	-0.00007	0.0000
	(0.0002)	(0.002)	(.0002)	(0.0000)
Δ(income per cap.)	0.0001	0.0002	0.0001	-0.0000
	(.00006)	(0.00007)	(0.00007)	(0.0000)
Δ(unemployment)	-0.094	-0.03	-0.056	-0.0001
	(.026)	(0.028)	(.032)	(0.0004)
Alcohol reg. controls	X	X	X	X
Year FE	X	X	X	X
$\hat{m}$	0.141	0.062	.403	
	(0.866)	(0.395)	(0.819)	
Sample size	1,487	1,389	937	1,487

Table 2: Estimating $\hat{m}$ following the strategy of CLK (2009) in the context of a linear model. As in CLK, we use the nominal excise tax rate divided by the average price of a case of beer from 1970 to 2003 as an IV for excise tax to eliminate tax-rate variation coming from inflation erosion. Next, we run the same regression in 3-year differences. Next, we run it only for states where food is exempt from sales-tax, to address concerns about whether consumers react differently to changes in the two taxes only because sales taxes apply to a broad set of goods. Finally, the last column addresses the concern that beer taxes may induce substitution with other alcoholic products, biasing the coefficient on excise tax relative to the one on sales tax. While in the log-log specification of CLK (2009) it seems to show that beer excise taxes have no discernable effect on the share of ethanol consumed from beer, we do find a significant effect. Standard errors in parentheses.

	Outcome variable: Number of cigarettes		
Specification	1	2	3
Excise Tax	-0.015	-0.015	-0.016
	(.004)	(.004)	(.004)
Sales Tax	-0.024	-0.02	-0.022
	(0.022)	(0.025)	(0.025)
Demographic controls	X	X	X
Econ. conditions controls		X	X
Income trend controls			X
State,year, and month FE	X	X	X
$\hat{m}$	1.57	1.33	1.37
	(1.65)	(1.83)	(1.82)
Sample size	274,138	274,138	274,138

Table 3: Estimating $\hat{m}$ based on the intensive response of cigarette consumption to sales taxes (not included in sticker price) and excise taxes (included in the sticker price). The specifications are a linearized version of the specifications in Goldin and Homonoff (2013). Standard errors in parentheses.

C Appedix Figures



(a) $p > p'$ (b) $p' > p$

Figure 5: A graphical illustration of Lemma 3. As long as demand is weakly decreasing, $dwl(p')$ cannot be smaller than $dwl(p)$ minus (plus) the orange rectangle.

Data Availability Statement: The data that support the findings of this study are openly available on the American Economic Association's website at http://doi.org/10.1257/aer.99.4.1145, reference number 10.1257/aer.99.4.1145, for the data in CLK (2009), and at http://doi.org/10.1257/pol.5.1.302, reference number 10.1257/pol.5.1.302, for the data in Goldin and Homonoff (2013).