Three Essays at the Intersection of Theory and Public Finance

by

Benjamin Glass

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2018

Doctoral Committee:

Professor James R. Hines Jr, Chair
Professor Tilman Borgers
Associate Professor Jeremy Fox, Rice University
Professor Joel Slemrod

Benjamin Glass

benglass@umich.edu

0000-0003-2422-5821

Dedication

To Meghan, the love of my life.

Acknowledgments

I have been blessed with an abundance of faculty and peers who have helped me selflessly throughout my time as a PhD student. I cannot offer sufficient gratitude for their assistance.

Thanks to Jim Hines for his invaluable support, both on this dissertation and in helping me find a job. It was through his encouragement that I discovered a love for public finance. I cannot imagine that I would have reached this milestone so soon without his encouragement and wisdom.

I must also thank two co-authors with whom I hope to continue working in the future. Joel Slemrod provided the initial inspiration for our co-authored paper, and he provided helpful guidance and support during a difficult job market. Giacomo Brusco worked tirelessly with me for years on our paper. I am lucky that he is both my colleague and my friend.

I owe my enthusiasm for general equilibrium theory to Tilman Borger and Alan Deardorff, both of whom have patiently discussed my work with me for years. Jeremy Fox agreed to be on my committee despite the difficulty of working hundreds of miles away. He also encouraged me at a difficult moment in my job search, for which I am thankful.

There are too many friends in the PhD program who warrant thanks for me to name them all. Yet Luis Alejos deserves acknowledgment for his comradery this past year. I have additionally benefitted from feedback and solidarity from Luis Baldomero, Will Boning, Owen

Nie, Tejaswi Velayudhan, Connor Cole, Kenichi Nagasawa, Parag Mahajan, Ari Binder, Aristos Hudson, and Meera Mahadevan.

Finally, two people deserve special recognition. My mom worked tirelessly as a single mother raising two difficult children. She encouraged me to pursue academia, for which I am very grateful. Also, asking her about her work as a doctor inspired the main idea for my third chapter. My fiancé Meghan supported me financially during my PhD program. She made sure I took care of myself during the program. She brought me joy during difficult moments. To both of you, I owe everything.

Preface

This dissertation is about rethinking old results in welfare economics. Each chapter revises previous understandings of normative analysis.

In the interest of getting bad news out of the way, the first chapter is the most negative result. Co-authored with Giacomo Brusco, it recalls the classic result that aggregate consumption data can identify the deadweight loss from a (sales) tax when that tax is fully salient. However, we demonstrate that, when one cannot rule out heterogeneity in the salience of the tax rate, aggregate demand data no longer point identifies deadweight loss. Imposing structure yields bounds on the possible values of deadweight loss, but these bounds empirically appear to yield substantial uncertainty in the calculation of deadweight loss.

The second chapter, co-authored with Joel Slemrod, provides a more positive result. It is well-understood that the burden of a sales tax does not depend on whether the buyer or the seller is responsible for mailing the check. We provide a far more general result in the same vein, where neither real economic activity nor welfare is affected by who must remit the taxes due to non-transfer activities. In this way, we reinforce a traditional result in public finance. However, that one can impose the responsibility to remit taxes arbitrarily without any welfare impacts suggests substantial latitude for the government to decide how to assign this responsibility based on non-welfarist considerations. Thus, our essay challenges previous results criticizing the consideration of horizontal equity in normative analysis.

The third chapter may be the most positive. In the standard competitive model, if firms have unboundedly increasing returns to scale, then no Pareto-efficient allocation that requires production is sustainable in a competitive equilibrium with transfers. However, I show that having agents make decisions about production, with an understanding that goods such as their own leisure cannot be purchased, allows for all Pareto-efficient allocations to be sustained in this variant of a competitive equilibrium with transfers. Yet this mechanism for attaining efficient allocations via price-taking behavior requires abandoning the classical dichotomy between consumption and production. Furthermore, now there is not generally a "profit" function that one can tax to obtain revenue without distorting behavior.

These three essays are written to be self-contained, but they touch upon similar elements. The first and third chapters both reinforce the difficulty of deadweight loss calculations. While the first chapter highlights the problem of heterogeneous tax salience, the third chapter implies that supply curves may also be subject to income effects that one should consider when performing welfare calculations. The first and second chapters imply that the government should be sensitive to how changing who remits taxes affects the salience of those taxes, even as the welfare impact of such a change in remittance responsibilities may be difficult to calculate. The second chapter anticipates the third by having agents make production and consumption decisions jointly.

Table of Contents

Dedication	ii
Acknowledgments	iii
Preface	v
List of Tables	viii
List of Figures	ix
Abstract	x
Chapter 1	1
Attending to Inattention: Identification of Deadweight Loss under Non-Salient Taxes	
Chapter 2	79
Tax Remittance Invariance	
Chapter 3	117
Doctors Need Sleep: A Theory of Optimal Production	

List of Tables

Table 1: Compensating variation and tax revenue for each type in each scenario. 20

Table 2: Lower and upper bounds of deadweight loss. 48

Table 3: Estimating $\widehat{m}$ with several sets of controls. 76

Table 4: Estimating $\widehat{m}$ following the strategy of CLK (2009) in the context of a linear model. 77

Table 5: Estimating $\widehat{m}$ based on the intensive response of cigarette consumption to sales taxes (not included in sticker price) and excise taxes (included in the sticker price). 78

Table 6: We report regression results directly from Table 1 of Rosen (1976). 78

List of Figures

Figures for Chapter 1

Figure 1.1: Consumer's decision under limited attention in Gabaix's model. 11

Figure 1.2: Welfare effects from the imposition of a non-salient tax. 15

Figure 1.3: A higher tax improves consumer welfare. 17

Figure 1.4: A graphical illustration of lemma 2. 26

Figure 1.5: A graphical illustration of Theorem 1. 28

Figure 1.6: A graphical illustration of Theorem 2. 32

Figure 1.7: Upper bound depending on the econometrician's choice of $\overline{m}$. 50

Figures for Chapter 2

Figure 2.1: An example of a connected network with functions κ and λ. 93

Figures for Chapter 3

Figure 3.1: The black curve delineates the boundary of the set R. 134

Abstract

This dissertation addresses three distinct tax questions. First, we address the identification of deadweight loss from non-salient taxes. Second, we generalize the canonical result that the impact of a tax does not depend on who has the legal responsibility to remit the tax payment. Finally, I extend the Walrasian equilibrium model to include unbounded increasing returns to scale production technology.

Taxes create deadweight loss by distorting consumer choice, so to the extent that consumers perceive taxes to be lower than they really are, deadweight loss declines. Deadweight loss is a convex function of perceived taxes, so its aggregate magnitude depends not only on the average tax misperception, but also on its heterogeneity among consumers. Aggregate data cannot reveal this heterogeneity, yet one can infer lower and upper bounds on deadweight loss relying solely on properties of aggregate demand. Sufficiently rich individual-level data permit identification of deadweight loss even with heterogeneous tax misperceptions. Under strong assumptions on the joint distribution of tax salience and preferences, survey data illustrate that tax salience heterogeneity can yield deadweight loss twice as large as one would calculate under the assumption of a homogeneous perceived price. Relaxing these assumptions, even slightly, yields much more destructive results: the unconstrained upper bound of deadweight loss is more than fifty times larger than the lower bound one would compute assuming homogeneous perceptions of price.

Conventional wisdom holds that the impact of a sales tax does not depend on whether the buyer or the seller remits the tax to the government. We extend this result to a general setting. Equilibrium non-transfer activity depends only on the total tax liability, regardless of who bears the statutory remittance liability. In a competitive setting, this result applies when all agents must pay each other for sales and purchases. However, changing the remittance obligation may transfer wealth lump-sum, depending on legal production and consumption rights. Thus, the remittance neutrality of taxes is a variation on Coase's Theorem.

A Walrasian equilibrium with production cannot exist when firms have unboundedly increasing returns to scale technology. If any price-taking firm chooses to produce at all, then it could achieve infinite profit by producing an infinite amount. Thus, the difficulty of increasing returns to scale in perfect competition results from firms acting without accounting for input supply constraints.

Chapter 1

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

by Giacomo Brusco and Benjamin Glass*

Abstract

Taxes create deadweight loss by distorting consumer choice, so to the extent that consumers perceive taxes to be lower than they really are, deadweight loss declines. Deadweight loss is a convex function of perceived taxes, so its aggregate magnitude depends not only on the average tax misperception, but also on its heterogeneity among consumers. Aggregate data cannot reveal this heterogeneity, yet one can infer lower and upper bounds on deadweight loss relying solely on properties of aggregate demand. Sufficiently rich individual-level data permit identification of deadweight loss even with heterogeneous tax misperceptions. Under strong assumptions on the joint distribution of tax salience and preferences, survey data illustrate that tax salience heterogeneity can yield deadweight loss twice as large as one would calculate under the assumption of a homogeneous perceived price. Without these assumptions, the unconstrained upper bound of deadweight loss is more than fifty times larger than the lower bound.

* University of Michigan. 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, William Boning, Luis Alejos, Tejaswi Velayudhan, Aristos Hudson, the participants of the Public Finance and Theory seminars at the economics department of the University of Michigan, and the participants of the 2017 NTA conference. We warmly thank all of them for their invaluable advice.

1. Introduction

Taxing a good results in a loss of economic efficiency whenever it distorts equilibrium behavior away from the Pareto optimum. If each person faced a different tax rate when buying a certain good, understanding the welfare effects of such a tax would require us to study not only aggregate demand responsiveness, but the demand responsiveness of every individual. Surely, 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, as noted in Chetty et al. (2009). Correctly assessing the welfare effects of taxation requires us to understand how each person reacts to the tax based on both their preferences and their perception of the tax. The possibility of heterogeneous attention intrinsically changes the way we should estimate deadweight loss. One can wildly miscalculate the magnitude of deadweight loss if they fail to account for the heterogeneous perception of taxes.

Suppose, e.g., that aggregate tax responsiveness was half of sticker price responsiveness. To a second order approximation, deadweight loss is quadratic in the perceived tax. If all consumers noticed only half of the tax, deadweight loss would be 25% of what it would be under full attention; if instead half of consumers did not notice the tax at all, while the other half noticed it perfectly, deadweight loss would be 50% of what it would be under full attention. Besides varying across individuals, attention may also endogenously co-vary with the tax rate.

Tax non-salience leaves consumers even worse off because they fail to avoid the tax by reducing consumption. Yet to what degree the tax does not distort consumer behavior, deadweight loss is reduced. In fact, we find that increasing an existing tax, while still increasing deadweight loss, can make consumers better off when it discourages consumption that would already be avoided if agents paid full attention.

Suppose the econometrician only has aggregate data: that is, data on aggregate consumption of a good across several markets, differentiated by time or geographic location. These data may indicate average responsiveness to taxes and sticker prices but provide no information on the heterogeneity of demand responsiveness to taxes, nor about its correlation with sticker price responsiveness. Thus, aggregate data cannot precisely identify deadweight loss. However, we can provide bounds for dead-weight loss even with aggregate data. These bounds are tight, in the sense that for each bound there is a distribution of salience under which that bound accurately describes deadweight loss.

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. Intuitively, heterogeneity in tax salience creates heterogeneity in perceived net-of-tax prices. This creates an allocative inefficiency across consumers: it is no longer true that the people who value the good the most are the ones who end up buying it. As the calculation with a representative consumer only accounts for inefficiency from *aggregate* foregone consumption due to the tax, it will underestimate excess burden.[1] 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.

We obtain an upper bound for deadweight loss by assuming 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. In addition to assuming tax salience is either zero or maximal, a distribution yielding the upper bound for deadweight loss assigns high tax salience to those agents whose preferences yield maximal deadweight loss from that agent relative to the change in consumption of that agent. In other words, this distribution allocates high tax salience to those agents who have more convex de-mand curves, keeping the aggregate change in quantity demanded constant. Finally, any agents with multiple optimal decisions at the perceived price consume the highest amount consistent with their preferences when they perceive low prices, whereas they consume the lowest amount consistent with their preferences when they perceive high prices. This last condition is need 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.

To obtain an understanding of the possible magnitude of the uncertainty in deadweight loss estimated with aggregate data, we perform an empirical exercise with experimental data

[1] While we came up with our lower bound independently, we thank Taubinsky and Rees-Jones (forthcoming) for inspiring this intuitive insight by making a related point as it pertained to their model. They note in proposition 7 of their appendix that deadweight loss can vary between lower and upper bounds when the choice set is binary.

from Taubinsky and Rees-Jones (forthcoming). For the good we observe, we estimate an average deadweight loss from sales taxes on that good across the country of between 0.38 cents and 20.79 cents per consumer. This means the upper bound of deadweight loss is around 55 times as large as the lower bound for deadweight loss. In fact, the upper bound for deadweight loss is substantially larger than if all agents were to perfectly account for the sales tax, in which case average deadweight loss would be 1.87 cents.

When able to observe consumption choices at the individual level with long panel data, one can point-identify deadweight loss. The easiest way to see this is to imagine one could observe the choice made by individual consumers for infinitely many periods, each time under a different tax and sticker price regime. Then, one would be able to infer sticker price and tax responsiveness for everyone, and so would be able to compute deadweight loss for each single agent, and thus for the entire population.

If we are willing to impose linear structure on the choice behavior of agents, then one need not observe each individual responsiveness to sticker prices and taxes: it will suffice to know their distribution in the population. For this, one need not follow the same agents across time; cross-sectional data with sufficiently rich variation in taxes and sticker prices will allow the identification of the distribution of responsiveness, which in turn would allow us to integrate for expected deadweight loss.[2]

An alternative to observing individual choices is to restrict the choice set that the agent is facing. Estimating random coefficients discrete choice models is common practice and can be

[2] See for instance Beran and Hall (1992), Beran, Feuerverger, and Hall (1996), and Hoderlein, Klemelä, and Mammen (2010), all of which would require full support of sticker prices and sales taxes.

achieved with aggregate data on market shares. The estimation of such a model allows the econometrician to then compute expected deadweight loss. This is because in the case of choices over a discrete set, aggregate data reveals the distribution of consumption on an individual level. For instance, if the good in question is a house, and assuming everyone either buys one house or does not buy a house at all, knowing the number of houses purchased and the size of the population of perspective buyers allows us to calculate the probability that an agent buys a house. However, the econometrician still requires linearity assumptions. In this context, these assumptions are equivalent to assuming that tax salience is independent of both the tax rate and the sticker price.

This paper complements a growing literature in public finance on non-salient taxes. Rosen (1976) does not find evidence of limited tax salience, but Chetty et al. (2009), Finkelstein (2009), Gallagher & Muehlegger (2011), Goldin & Homonoff (2013), and Taubinsky & Rees-Jones (forthcoming) all find strong evidence of dramatically limited tax salience. For instance, Chetty et al. (2009) provide a summary estimate that agents perceive six percent of sales tax variation using a log-log specification. We estimate a linear specification using their data and find a summary estimate of 27 percent.

Recent theoretical papers on the efficiency implications of non-salient taxes include Chetty et al. (2007), Farhi and Gabaix (2015), and Taubinsky and Rees-Jones (forthcoming). Our theoretical model comes from Gabaix (2014), which differs slightly from Chetty et al. (2009) in how it handles income effects. In our model, a sales tax of any positive salience creates deadweight loss.[3] In contrast, Goldin (2015) demonstrates that a model developed by Chetty et

[3] This claim assumes an undistorted economy without perfectly inelastic supply or demand

al. (2009) generically yields zero deadweight loss for some positive sales tax on a normal good.[4] We show that the Gabaix (2014) model provides a general description of behavior when agents have convex preferences and proceed to use it for welfare analysis.

Taubinsky and Rees-Jones (forthcoming) is most like this paper in spirit. They make similar points about the inability to identify deadweight loss with aggregate data due to the role played by heterogeneous attention. They also find lower and upper bounds for a second order approximation to deadweight loss with a binary choice set. We find similar bounds without imposing restrictions on the choice set.

This paper proceeds as follows. In section 2, we develop the decision theory model and theoretically derive deadweight loss. In section 3, we illustrate how heterogeneity in attention prevents identification of deadweight loss with aggregate continuous choice data. We then show how we can still identify lower and upper bounds to deadweight loss under weak assumptions. We provide positive point identification results in section 4 using individual-level or binary choice data.[5] We perform an empirical calculation in section 5 before concluding in section 6. We relegate proofs and details of empirical work to the appendix.

2. Theoretical Derivation of Deadweight Loss

This section describes the theoretical model and results that underlie the rest of this paper. The main modeling challenge in dealing with misperceived prices is to allow for the

[4] This result uses a two-good framework in which the untaxed good absorbs any budget shortfalls. In contrast, Goldin's (2015) result relies on using the ratio of uncompensated demand responses to tax and sticker price variation. In our model, tax salience corresponds to a ratio of compensated demand responses.

[5] It may appear strange that Taubinsky and Rees-Jones (forthcoming) find a negative identification result with binary choice data, whereas we find a positive identification result. The distinction comes from whether one assumes that tax salience does not depend on the sticker price or tax rate.

misperception of prices while keeping agents solvent. Chetty et al. (2009) get around this issue by having a single good “absorb'' all optimization mistakes. Gabaix (2014), instead, has agents conjecture themselves a certain income such that they end up consuming on their true budget constraint when presented with the relative prices they perceive. We begin with a general model of decision-making problem for a single agent that encompasses both the Chetty et al. (2007, 2009) and Gabaix (2014). The choice behavior of the agent is modeled in a general manner, so that the agent correctly optimizes for the goods without non-salient taxes, but has an arbitrary (continuous) consumption function for the taxed good. We then show that the choice behavior from such a model can be represented by the model from Gabaix (2014), which dramatically simplifies exposition. In the process, we define compensating variation and analytically characterize it using the Gabaix (2014) model. Finally, we aggregate over agents and account for the change in government revenue to derive aggregate deadweight loss due to the tax.

2.1 Choices under Misperceived Taxes

We consider a decision problem over two goods, one with a non-salient sales tax. We generalize to multiple taxed and non-taxed goods in the appendix. The agent has closed consumption set $X = X^T \times \mathbb{R}_+ \subseteq \mathbb{R}_+^2$. The agent has 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 the sticker price of the taxed and non-taxed good, respectively, $\tau \in \mathbb{R}$ is the sales tax on the taxed good, and W is the income of the agent.[6] 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 continuous and strictly monotonic preferences $\succcurlyeq$ with continuous utility representation $u(q, q^{NT})$, where q denotes generic consumption of the taxed good. The choice vector function $\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W) = \left(q(\bar{p}, p^{NT}, \tau, W), q^{NT}(\bar{p}, p^{NT}, \tau, W)\right) \in X$ satisfies two requirements. One, the agent always spends all available income:

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

In the appendix, we generalize so that agents optimally choose q^{NT} given their choice of q.

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

$$\boldsymbol{q}(\bar{p}, p^{NT}, 0, W) \in \underset{\bar{p}*q+p^{NT}*q^{NT} \leq W}{\operatorname{argmax}} u(q, q^{NT})$$

We 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:[7]

$$\Delta CS \equiv \inf \left\{\Delta W \mid u\big(\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W + \Delta W)\big) \geq u\big(\boldsymbol{q}(\bar{p}, p^{NT}, 0, W)\big)\right\}$$

In words, the change in consumer surplus is the greatest lower bound of the amount of money we must provide the agent so that the agent achieves the utility from before the imposition of

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

[7] We impose assumptions in subsection 2.2 that would allow us to define compensating variation using a minimum, rather than an infimum.

the tax. One can verify that compensating variation is always non-negative for any non-negative tax when preferences are locally non-satiated.[8]

2.2 Gabaix Representation of Choice Behavior

We now provide sufficient conditions that permit one to write the above model in the framework of Gabaix (2014) without loss of generality. In this model, the agent perceives price p^s for the non-saliently taxed good, correctly perceives price p^{NT} for the non-taxed goods, and conjectures an income W^s so that the agent's true budget constraint is satisfied while the agent optimizes subject to the perceived budget constraint. Formally:[9]

$$\boldsymbol{q}(\bar{p}, p^{NT}, \tau, W + \Delta W) \in \underset{p^s q + p^{NT} q^{NT} \leq W^s}{\operatorname{argmax}} u(q, q^{NT})$$

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

To find a Gabaix representation for a choice, we want to find a budget line through the consumption bundle such that all strictly preferred bundles lie strictly above this line. This is analogous to finding an equilibrium to a single-agent endowment economy. As in the Second Welfare Theorem, we require some sense of continuous convexity of preferences. However, we do not want to assume that the choice set is convex, to later allow for treatment of discrete choice sets. Instead, we impose convexity via the utility representation.[10]

[8] For any $\Delta W < 0$, $(p + \tau, p^{NT}) * \boldsymbol{q}(\bar{p}, p^{NT}, \tau, W + \Delta W) = W + \Delta W < W = (p + \tau, p^{NT}) * \boldsymbol{q}(\bar{p}, p^{NT}, 0, W)$, implying that $u(\boldsymbol{q}(\bar{\boldsymbol{p}}, \tau, W + \Delta W)) < u(\boldsymbol{q}(\bar{\boldsymbol{p}}, 0, W))$.

[9] Technically, for the subjective price p^s, Gabaix (2014) also wanted to choose W^s maximally amongst all values of W^s that would satisfy these two equations. The idea is for the agent to conjecture income to optimally satisfy the true budget constraint. This addition is without loss of generality if the taxed good is weakly normal, since then there is at most one conjectured income that satisfies the true budget constraint.

[10] In the appendix, we generalize this result with a description of properties of the choice set and preferences. The key idea is still that there is an open and convex set containing all the bundles strictly preferred to $\boldsymbol{q}(p, p^{NT}, \tau, W)$.

Figure 1.1

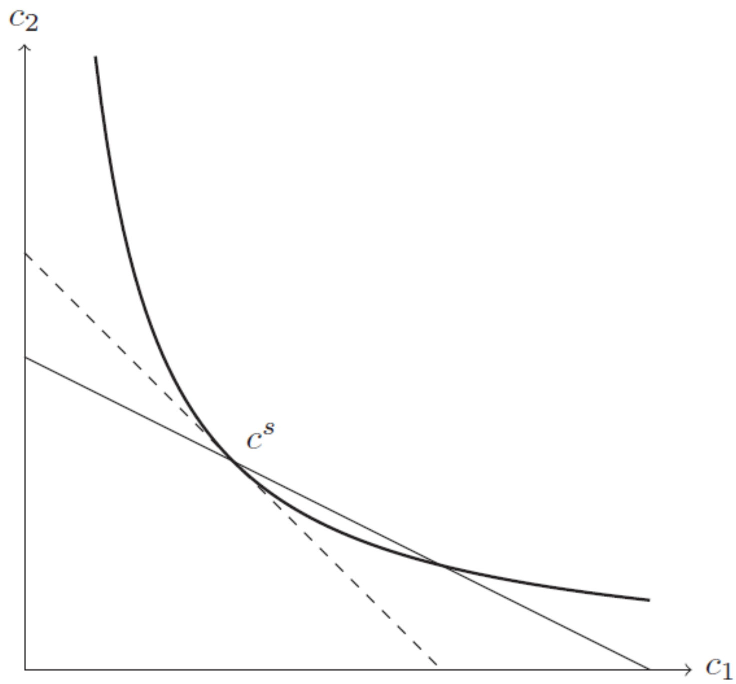


Figure 1.1: Consumer's decision under limited attention in Gabaix's model. The solid line represents the consumer's real budget constraint. For any chosen point c^s on a convex indifference curve, one can always find a supporting hyperplane, represented by the dotted line, that rationalizes the point as a budget line. Source for image: Gabaix (2014).

Proposition 1: 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:

$$\boldsymbol{q}(p, p^{NT}, \tau, W) \in \underset{p^s q + p^{NT} q^{NT} \leq W^s}{\operatorname{argmax}} u\left(q, q^{NT}\right)$$

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

The proof idea is that the set of points $\left\{\boldsymbol{q}' \in \mathbb{R}^2 \middle| \, u(\boldsymbol{q}') > u\left(\boldsymbol{q}(p, p^{NT}, \tau, W)\right)\right\}$ is open and convex. Since it's convex, there is a budget line separating it from $\boldsymbol{q}(p, p^{NT}, \tau, W)$. Since it is open, that set can never touch this budget line. Finally, we can always shift the budget line down if need be so that it goes through the point $\boldsymbol{q}(p, p^{NT}, \tau, W)$.

Note that proposition 1 does not rule out alternate explanations for consumer behavior. For instance, Chetty, Looney, and Kroft (2009) have a model in which $W^s = W$. The consumer first buys the taxed good knowing the available total income, then re-optimizes upon discovering how little income remains after purchasing the taxed good. Instead, the proposition notes that any such model satisfying minimal conditions is observationally equivalent to the Gabaix model. Intuitively, the model states that the agent doesn't notice some fraction of the tax-inclusive price, instead simply figuring that the extra amount spent on the taxed good was never in the bank in the first place. This story may appear unlikely, but one need not take it literally. Instead, it is a representation of consumer behavior from a general model. Still, the intuition helps us understand our characterization of the compensating variation due to the tax. In the following proposition, we introduce terminology with some arguments suppressed for ease of reference.

Proposition 2: Let $e(p)$ and $h(p)$ respectively denote the expenditure and compensated demand functions for the taxed good at price p for the taxed good and price p^{NT} for the other good, so that the agent is minimally compensated to achieve utility of a least $u\big(\boldsymbol{q}(p, p^{NT}, 0, W)\big)$.[11] Then compensating variation due to the tax satisfies:

$$\Delta CS = (\bar{p} + \tau - p^s)\; q(\bar{p}, p^{NT}, \tau, W + \Delta CS) + e(p^s) - e(\bar{p}) \qquad (1)$$

where p^s is a Gabaix representation of the perceived price with the tax.[12]

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

---

[11] Formally, $e(p) \equiv \min\{W' \mid u\big(d(p, p^{NT}, W'), d^{NT}(p, p^{NT}, W)\big) \geq u\big(\boldsymbol{q}(p, p^{NT}, 0, W)\big)\}$, which is well-defined because u is continuous and the choice set is compact.

[12] The same choice model may have multiple Gabaix representations, particularly if the taxed good is inferior.

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

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:

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

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

Rearranging and again using local non-satiation yields:

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

This result has a natural interpretation. The difference in expenditure functions $e(p^s) - e(\bar{p})$ represents the amount the consumer would have to be compensated if the tax-inclusive price were *actually* p^s. But instead, the agent is paying an extra $(\bar{p} + \tau - p^s)$ per unit of the taxed good consumed, and so must be compensated for that "lost" income. Furthermore, we can define $h(p)$ as the compensated demand with income $e(p)$. The expenditure function is concave, and so has well-defined derivatives almost everywhere. By Shephard's Lemma, these derivatives are Hicksian demand. By the Fundamental Theorem of Calculus:

$$e(p^s) - e(\bar{p}) = \int_{\bar{p}}^{p^s} h(p)dp$$

Thus, if $h(p^s)$ is well-defined, we can express the change in consumer surplus as:

$$\Delta CS = [\bar{p} + \tau - p^s] * h(p^s) + \int_{\bar{p}}^{p^s} h(p)dp \qquad (2)$$

From the representation of deadweight loss, we can immediately derive a couple of insights. First, suppose for a moment that $\tau > 0$ and $p^s \in [\bar{p}, \bar{p} + \tau]$, so that the agent does not fully notice the increase in the tax-inclusive price. We can then confirm that the non-salience of the tax weakly exacerbates the loss of consumer surplus:

$$\Delta CS = [\bar{p} + \tau - p^s] * h(p^s) + \int_{\bar{p}}^{p^s} h(p)dp \geq \int_{p^s}^{\bar{p}+\tau} h(p)dp + \int_{\bar{p}}^{p^s} h(p)dp = \int_{\bar{p}}^{\bar{p}+\tau} h(p)dp$$

This reflects the fact that consumers only partially respond to the tax. If they completely understood how the tax was affected the tax-inclusive price, they would protect themselves from this tax burden by reducing consumption of the taxed good. As they fail to fully account for the tax, they end up worse off.

Second, empirical work often uses linear demand functions. We will also consider linear choice functions in section 3. As in the standard decision-making model, the calculation of the change in consumer surplus calculated using sticker price and tax derivatives naively as if the choice function were linear in these arguments is a second order approximation to the true change in consumer surplus. We now demonstrate this claim formally.

Assume that h is continuously differentiable with respect to its own price. Assume p^s is continuously differentiable with respect to τ, so that we can define $m \equiv \frac{\partial p^s}{\partial \tau}|_{(\bar{p},0)}$.[13] Taking a second order approximation around $\tau = 0$ then yields:[14]

$$\Delta CS \approx h(\bar{p})\,\tau + \left[(1-m)m\,\frac{\partial h}{\partial p}|_{\bar{p}} + m\,\frac{\partial h}{\partial p}|_{\bar{p}}\right]\tau^2/2$$

[13] Formally, the claim is that there is a Gabaix representation that has well-defined $\frac{\partial p^s}{\partial \tau}$ for a neighborhood of tax rates around zero, where the derivative is taken while the consumer is being compensated. 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 $\frac{\partial p^s}{\partial \tau}|_{\tau=0} = 0$ and $\frac{\partial \Delta CS}{\partial \tau} = -\frac{\frac{\partial q}{\partial \tau}}{\frac{\partial q}{\partial W}}$.

[14] This derivation appears to generalize proposition 22 from Gabaix (2014).

Figure 1.2

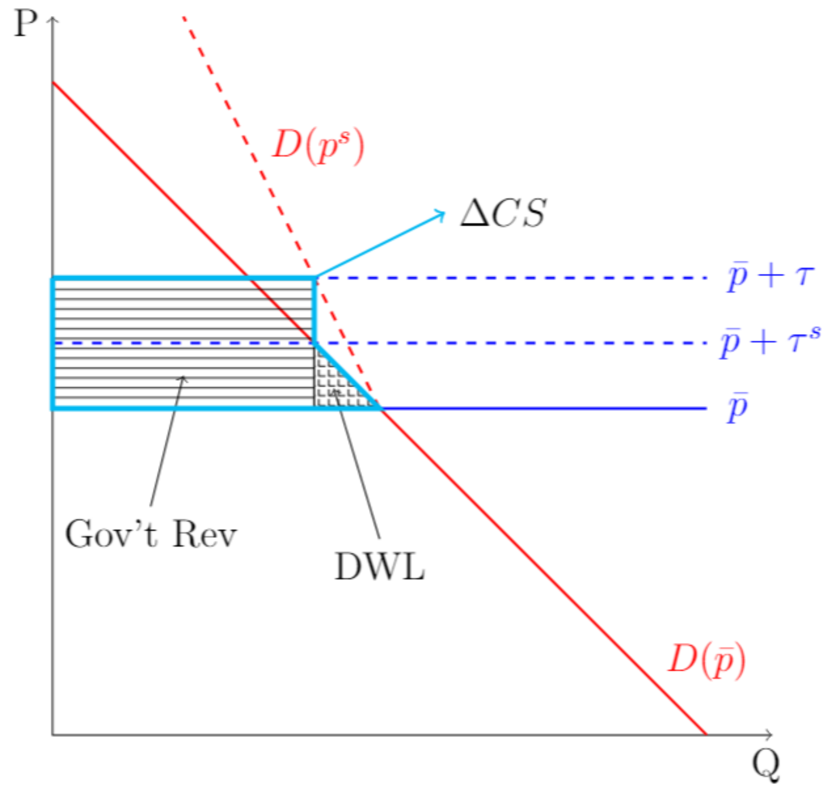


Figure 1.2: Welfare effects from the imposition of a non-salient tax.

Rearranging yields:

$$\Delta CS \approx \left[h(\bar{p})\,\tau + \left(m\,\frac{\partial h}{\partial p}|_{\bar{p}}\right)\tau^2/2\right] + (1-m)m\,\frac{\partial h}{\partial p}|_{\bar{p}}\,\tau^2/2$$

The first term in this calculation, $h(\bar{p})\,\tau + \left(m\,\frac{\partial h}{\partial p}|_{\bar{p}}\right)\frac{\tau^2}{2}$, is the change in consumer surplus one would naively expect from using tax (compensated) responsiveness as if the agent were fully attentive to the tax. But this evaluation does not account for the discrepancy between the tax rate and the true marginal value of the good to the consumer. The second term, $(1-m)m\,\frac{\partial h}{\partial p}|_{\bar{p}}\,\frac{\tau^2}{2}$, reflects how the agent effectively loses income to the lump-sum portion of the tax $(\bar{p} + \tau - p^s)$, so that further increasing the tax motivates the consumer to reduce consumption and so mitigate the income lost to non-salient taxation. More formally, note that the amount of income the agent loses to the tax without noticing is:

$$(1-m)\, q^s\, \tau$$

When the tax marginally increases from zero, it causes demand to decrease marginally by $m\, \frac{\partial h}{\partial p}$. Integrating $(1-m)m\, \frac{\partial h}{\partial p}\, \tau$ from zero to τ yields the term $(1-m)m\, \frac{\partial h}{\partial p}\frac{\tau^2}{2}$ (when ignoring the curvature of Hicksian demand as in the second order approximation). Thus, the second term is an approximation around $\tau = 0$ of the income no longer "misplaced" due to the tax. It is an *internality* in that it is a benefit to the consumer for which the consumer does not account. This measure of the internality is generically substantial. In fact, it can overwhelm the direct harm from the tax at the margin. If the non-salient tax is already sufficiently large, then it is possible for a marginal increase in the tax to increase consumer welfare, as illustrated in Figure 3.[15]

To see this result formally, suppose for a moment that demand d were linear and without income effects. Still suppressing the price of the other good, the loss of consumer surplus is then:

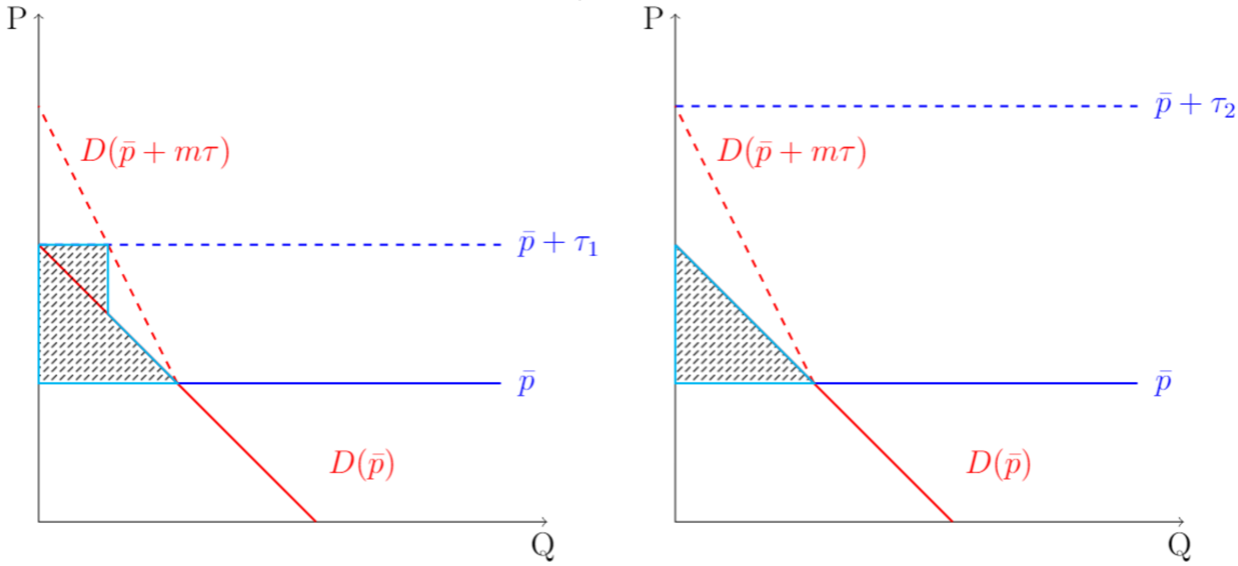
$$\Delta CS = d(\bar{p}, W)\, \tau + m\, \frac{\partial d}{\partial p}\frac{\tau^2}{2} + (1-m)\, m\, \frac{\partial d}{\partial p}\frac{\tau^2}{2}$$

The rate of change of the loss of consumer surplus is:

$$\frac{\partial \Delta CS}{\partial \tau} = d(\bar{p}, W) + m\, \frac{\partial d}{\partial p}\, \tau + (1-m)\, m\, \frac{\partial d}{\partial p}\, \tau$$

[15] It may appear strange that we were previously discussing a second order approximation around $\tau = 0$, but now discuss tax changes "sufficiently large". The possibility of a further tax increase benefitting a consumer holds without reference to any functional form, but we need not refer to second order approximations if q is linear with respect to $\bar{p}$ and τ.

Figure 1.3



(a) Under τ_1 the consumer loses more than total CS (b) Under τ_2 the consumer loses exactly total CS

Figure 1.3: A higher tax improves consumer welfare. The consumer is better off in (b) than (a) from the higher tax rate reducing consumption.

Linear demand implies that quantity is positive whenever $d(\bar{p}, W) + m\,\frac{\partial d}{\partial \tau}\,\tau > 0$. So, from the Law of Demand and assuming $m \in (0,1)$:

$$d(\bar{p} + m\tau, W) = 0 \Rightarrow \frac{\partial \Delta CS}{\partial \tau} < 0$$

From continuity, we can conclude that the loss of consumer surplus decreases whenever the tax is sufficiently high such that consumption is sufficiently small (but positive). In other words, once the tax is sufficiently large to make consumption sufficiently small, increasing the tax further benefits the consumer.

Given the potentially large and qualitatively important correction to welfare due to the internality from the tax, we do not recommend ignoring it.[16] Instead, we calculate deadweight loss while accounting for this internality. Deadweight loss from the tax is the difference

[16] One might dismiss the internality as second-order. But even with fully salient taxation, all efficiency loss is second-order.

between the income necessary to compensate the consumer from the tax and the revenue that would be raised from the tax (with the consumer compensated):[17]

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

Note that one calculates deadweight loss as if with perceived tax $\tau^s \equiv p^s - \bar{p}$. With well-defined and sufficiently smooth compensated demand:

$$dwl = [e(p^s) - e(\bar{p})] - [p^s - \bar{p}]\, h(p^s) = \int_{\bar{p}}^{p^s} h(p)dp - [p^s - \bar{p}]\, h(p^s) \qquad (3)$$

Taking a second order approximation around $p^s = \bar{p}$ yields:

$$dwl \approx -\frac{1}{2} m^2 \,\frac{\partial h}{\partial p}\, \tau^2$$

2.3 Aggregate Deadweight Loss from a Non-Salient Tax

In this subsection, we derive aggregate deadweight loss and discuss some of its properties. Our principle message in this subsection is that, in a model of misperceived prices, different joint distributions of perceived prices and preferences can yield the same aggregate demand but result in extremely different deadweight loss.

We start with a generic environment with a non-taxed good, priced at p^{NT}, and a taxed good, with sticker price $\bar{p}$ and a (per-unit) tax τ. Let $i \in \mathcal{I}$ index consumers. Each consumer is characterized by her perception of the price of the taxed good p_i^s and type θ_i for preferences $\succcurlyeq_{\theta_i}$ & income W_{θ_i}. In addition, consumers have tie-breaking parameters ζ_i, the importance of which we clarify momentarily. The consumer-specific parameters have joint distribution $F^*_{p^s,\theta,\zeta}$. Each agent has a choice function for the taxed good, satisfying:

[17] We maintain the convention that deadweight loss is generically positive.

$$q(\bar{p}, p^{NT}, \tau, W_{\theta_i}; \theta_i, \zeta_i) \in \mathcal{Q}_{p_i^s,\theta_i} \equiv \left\{ \begin{array}{c} q \mid \exists\ q^{NT} \in X^{NT}: p_i^s * q + p^{NT} * q^{NT} \leq W_i^s, \\ (q, q^{NT}) \succcurlyeq_{\theta_i} \boldsymbol{q}' \ \forall\, \boldsymbol{q}' \in X: (p_i^s, p^{NT}) * \boldsymbol{q}' \leq W^s \end{array} \right\}$$

where W_i^s is endogenously determined as in the Gabaix model, with corresponding expenditure function $e(p_i^s; \theta_i)$.

The parameter ζ_i breaks ties among bundles that could have been chosen: choices do not necessarily reflect true preferences when agents misperceive prices, and agents might appear indifferent between choices that do not yield the same (ex-post) utility. This sharply differs from the neo-classical model, in which the actual choice that one selects among indifferent bundles has no impact on consumer surplus.

For example, suppose agents face the problem of spending a fixed income $W > 2$ on a binary taxed good $q \in \{0,1\}$, with a sticker price $\bar{p} = 1$ & a tax of $\tau = 1$, and a non-taxed good $q^{NT} \in \mathbb{R}_+$, priced at $p^{NT} = 1$. There are two types of agents. Half of agents maximize utility $u_1(\boldsymbol{q}) = q + q^{NT}$, and perceive $p_1^s = 1$, so they fail to see the tax and are indifferent about buying q both before and after the imposition of the tax. The other half maximize $u_2(\boldsymbol{q}) = 2q + q^{NT}$, and $p_2^s = 2$, so agents of type 2 perfectly notice the tax, are indifferent to buying the good after the tax, and always buy q without the tax.

Let us suppose for concreteness that everyone buys q before the tax is imposed.[18] Consider two possible scenarios occurring after the imposition of the tax. In the first scenario all agents of type 1 buy q and no agents of type 2 buy q. In the second scenario, no agents of type 1 buy q and all agents of type 2 buy q. We work out the details in the appendix, but we summarize results for aggregate deadweight loss in table (1), so that one can check that

[18] The punchline of this example does not depend on this assumption, as either way consumers of type one obtain utility W before the tax is imposed.

Table 1

	Type	q pre-tax	$\boldsymbol{q}$ post-tax	W^s post-tax	ΔCS	Tax Revenue	Total DWL
Scenario 1	Type 1	$(1, W-1)$	$(1, W-2)$	$W-1$	1	1	0.5
	Type 2	$(1, W-1)$	$(0, W)$	W	1	0	
Scenario 2	Type 1	$(1, W-1)$	$(0, W)$	W	0	0	0
	Type 2	$(1, W-1)$	$(1, W-2)$	W	1	1	

Table 1: Compensating variation and tax revenue for each type in each scenario. Total deadweight loss is the weighted average of compensating variation in excess of tax revenue.

deadweight loss will differ depending on the final choices of agents who are "indifferent". This is because when agents misperceive prices, they can appear indifferent between choices that they value differently: in the example, agents of type 1 are happier when they don't buy the good (as if they perceived the tax), but in the model their final choice depends on whether they either conjecture an income $W_1^s = W$ while choosing $\boldsymbol{q} = (0, W)$ or conjecture $W_1^s = W - 1$ while choosing $\boldsymbol{q} = (1, W-2)$. Which choice they make is governed by the parameter ζ_i.

Note that, although the Gabaix model does differ from the neoclassical consumption model in several ways, the introduction of this notation does not challenge our intuitive understanding of compensated demand. Consider for instance two values l and h such that for any p_i^s and θ_i:

$$[q(p_i^s; \theta_i, l), q(p_i^s; \theta_i, h)] \supseteq \mathcal{Q}_{p_i^s, \theta_i}$$

So $q(p_i^s; \theta_i, l)$ and $q(p_i^s; \theta_i, h)$ are the smallest and largest amounts, respectively, of the taxed good that the agent could choose to consume. With that in mind, lemma 1 shows that the Law of Compensated Demand holds in this framework.

Lemma 1: For any agent i with type θ_i and any two prices p and p':

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

We prove this result in the appendix.

For the rest of this section, we assume well-defined and sufficiently smooth compensated demand so that $Q_{p_i^s,\theta_i}$ is always single-valued. Suppressing ζ, deadweight loss with complete pass-through takes the form:

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

In line with most analysis of deadweight loss, we can now consider a second-order approximation, which allows us to characterize our object of interest in terms of first derivatives:

$$DWL \approx -\frac{1}{2}\int_{p_i^s,\theta_i} m_i^2 \frac{\partial h}{\partial p}(\bar{p};\theta_i)\, dF^*_{p^s,\theta}(p_i^s,\theta_i)\, \tau^2$$

Note that neither aggregate price responsiveness $\int_{p_i^s,\theta_i} \frac{\partial h}{\partial p}(\bar{p};\theta_i)\, dF^*_{p^s,\theta}(p_i^s,\theta_i)$ nor aggregate tax responsiveness $\int_{p_i^s,\theta_i} m_i \frac{\partial h}{\partial p}(\bar{p};\theta_i)\, dF^*_{p^s,\theta}(p_i^s,\theta_i)$ are sufficient statistics for deadweight loss as a function of the tax rate. This illustrates the challenge of calculating deadweight loss from observable data.

Finally, we wish to consider the impact on aggregate deadweight loss of allocative inefficiency. So far, we have considered perfectly elastic supply, so that the post-tax sticker price remains unchanged at $\bar{p}$. In the case of an arbitrary differentiable aggregate supply function Q^{supply}, deadweight loss has second order approximation:[19]

$$DWL \approx -\frac{1}{2}\left[\int_{m_i,\theta_i} m_i^2\ \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} - \frac{\left(\int_{m_i,\theta_i} m_i\ \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} dF_{m,\theta}\right)^2}{\int_{m_i,\theta_i} \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} dF_{m,\theta} - \frac{\partial Q^{supply}}{\partial p}}\right]\tau^2$$

[19] We formally demonstrate this result with a finite population in the appendix. To make this result general, one need only assume that the price derivative of compensated demand is uniformly bounded on the support of (p^s,θ).

The additional term from incomplete pass-through is entirely determined by aggregate (compensated) tax responsiveness, aggregate (compensated) price responsiveness, and aggregate price responsiveness of supply. So, with appropriate aggregate data, the problem of calculating deadweight loss in a general setting reduces to the problem of calculating deadweight loss with complete pass-through. This motivates our principal concern with determining deadweight loss from observable data while assuming complete pass-through.

However, consider for a moment the case of perfectly inelastic supply, i.e. $\frac{\partial Q^{supply}}{\partial p} = 0$. Then:

$$DWL \approx -\frac{1}{2}\left[\int_{m_i,\theta_i} m_i^2 \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} - \frac{\left(\int_{m_i,\theta_i} m_i \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} dF_{m,\theta}\right)^2}{\int_{m_i,\theta_i} \frac{\partial h}{\partial p}|_{(\bar{p};\theta_i)} dF_{m,\theta}}\right] \geq 0$$

Note that deadweight loss is zero when $m_i = m \; \forall i$.[20] The imposition of the tax forces the sticker price to decrease so that aggregate consumption remains constant. If there is heterogeneity in attention, then different consumers perceive different prices even while facing the same tax. Some agents perceive a higher price due to the tax, while others perceive a lower price as a result. Those who perceive a higher price reduce their consumption, with that foregone consumption going to those perceiving a lower price. This creates allocative inefficiency, as units of the good are not necessarily going in the hands of the people who value them the most.[21]

[20] Technically, the previous expression only shows that the second order approximation for deadweight loss is zero in the case of homogeneous attention with perfectly inelastic supply. We demonstrate in the appendix that deadweight loss is precisely zero in this case.

[21] While we believe this result to be novel as it pertains to our model, we credit Taubinsky and Rees-Jones (forthcoming) for the insight that taxation still generically yields allocative inefficiency even when supply is perfectly inelastic.

Now that we have looked at how deadweight loss works in aggregate when taxes are misperceived, we next tackle the problem of identifying it with data. Section 3 deals with identification in the case where the econometrician can only observe aggregate demand across markets; we will find that while point-identification is impossible, one can still bound deadweight loss. In section 4 we instead deal with the case where the econometrician can observe demand across individuals and discuss under which assumptions one might be able to point-identify deadweight loss.

3. Non-Identification with Aggregate Continuous Choice Data

This section discusses to what degree one can infer deadweight loss from aggregate choice data. Most of the results from this section apply generally. For simplicity, we will assume that the econometrician has already determined the distribution of preference types. In this case, the econometrician can provide a lower bound for deadweight loss by assuming that all agents perceive the same tax-inclusive price, i.e. assume there is no attention heterogeneity. Alternatively, one can derive an upper bound for deadweight loss by supposing maximal attention heterogeneity. Since the data do not reveal the variance in tax salience, one cannot point identify deadweight loss from aggregate data.[22] When the consumption function is linear in the sticker price and tax rate, deadweight loss can take on any value between the upper and lower bounds. The results in this section are described as if all agents face the same sales tax, but none of our results depend on that assumption.

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 requires strong

[22] This claim holds generically, with an exception if the tax does not alter aggregate consumption.

restrictions on income effects to achieve identification with aggregate data. Suppressing income 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 $q(p_i^s; \theta_i, \zeta_i)$.[23] To ensure integrability, we assume the econometrician knows that $F_{p^s}^*$ has support bounded above 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:

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

The problem of identification is to find conditions for which any joint distribution $F_{p^s,\theta,\zeta}$ of (p^s, θ, ζ) satisfying these conditions (as a function of observable variables) such that:

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

yields the same value for deadweight loss:

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

[23]All results follow if one reinterprets $q(p_i^s; \theta_i, \zeta_i)$ as the compensated choice of agent i.

The main message of this section will be the failure of such a result to obtain under plausible restrictions arising from aggregate data alone.

We will use the following lemma to derive bounds on the possible values of deadweight loss:

Lemma 2: For any agent i with type θ_i and any two pairs (p, ζ_i) & (p', ζ_i'):

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

This lemma comes entirely from the compensated law of demand (CLD), and can be confirmed with a simple graph, as we show in figure 1.4. This lemma indicates the convexity of deadweight loss with respect to the perceived price.

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:

$$\begin{aligned} e(p') - e(p) &= (p', p^{NT}) * \boldsymbol{q}' - (p, p^{NT}) * \boldsymbol{q} \geq (p', p^{NT}) * \boldsymbol{q}' - (p, p^{NT}) * \boldsymbol{q}' \\ &= [p' - p] * q(p'; \theta_i, \zeta_i') \end{aligned}$$

Plugging in yields:

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

Figure 1.4

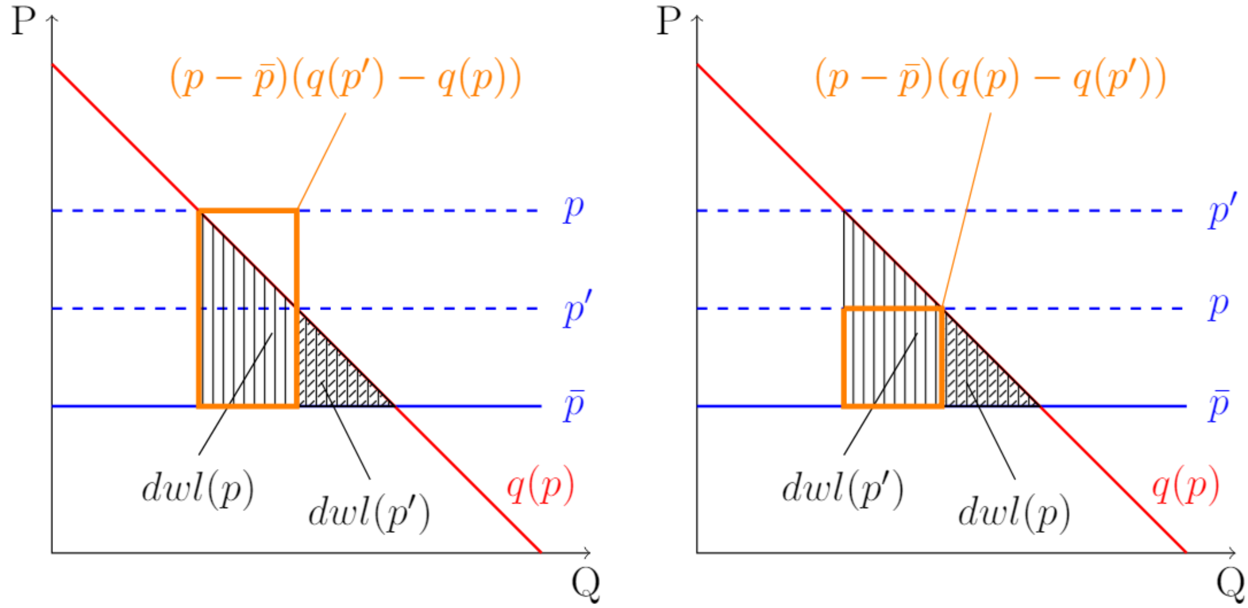


Figure 1.4: A graphical illustration of lemma 2. Since demand is weakly decreasing, dwl(p′) cannot be smaller than dwl(p) minus (plus) the orange rectangle.

Finally, we impose regularity conditions 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 support of (p^s,θ,ζ) on a finite set satisfy the above conditions.

3.1 Lower Bound on Deadweight Loss

Even though we cannot point identify deadweight loss with aggregate data, we can identify upper and lower bounds. For the lower bound, consider arbitrary $\bar{p}$, p^{NT}, and τ. For arbitrary $F_{p^s,\theta,\zeta}$ that could generate the data, we can choose a price $\hat{p}^s$ that could also rationalize the data if perceived by everyone.

Proposition 3: For any $F_{p^s,\theta,\zeta}$ that yields integrable aggregate demand, there is a scalar $\hat{p}^s$ such that for some distribution $F'_{\theta,\zeta}$ such that $F'_\theta = F_\theta$:

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

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:

$$\int_{p_i^s,\zeta_i} q(\hat{p}^s;\theta_i,\zeta_i)dF_{\theta,\zeta}(\theta_i,\zeta_i) = \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s;\theta_i,\zeta_i) \tag{4}$$

Then the following inequality obtains:

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

The intuition behind this result is that introducing heterogeneity in perceived prices facilitates gains from trade by having agents trade with each other after making their consumption decisions. If one person perceives a higher price than another, then the two agents will have different marginal valuations of the good. If they could exchange with each other, the one who perceived the higher price could purchase some of the good from the other

Figure 1.5

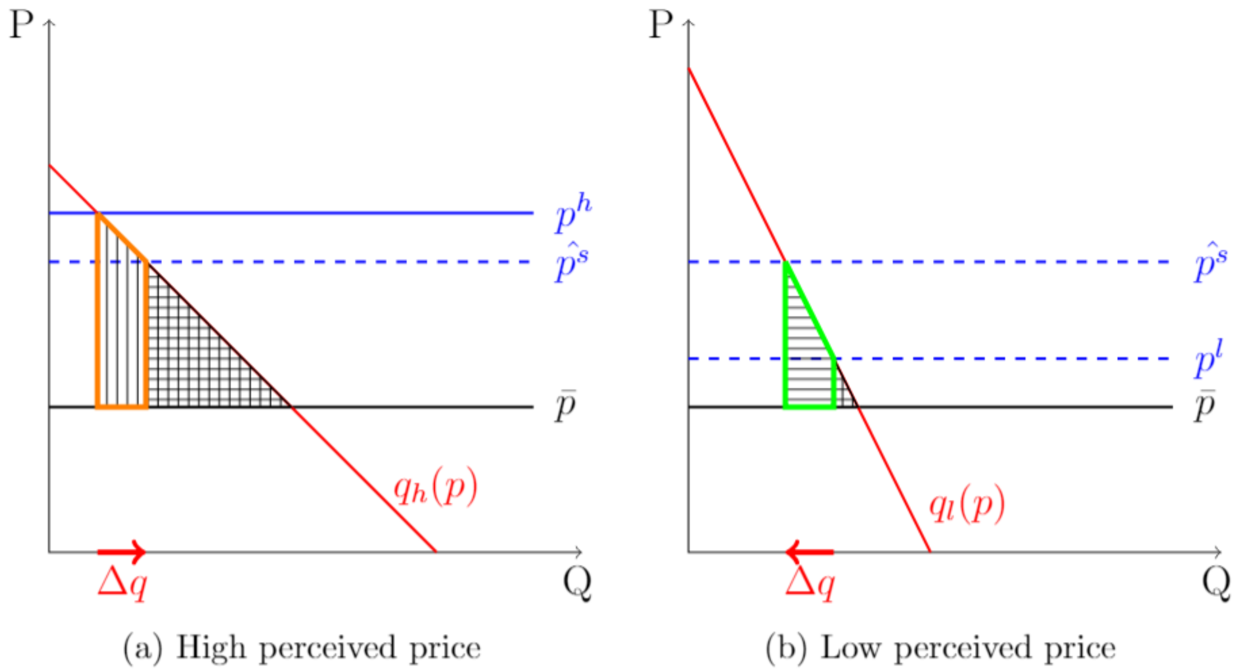


Figure 1.5: A graphical illustration of Theorem 1. When one picks $\hat{p}^s$ as to make the change in demand equal for the consumer in (a) and in (b), the (orange) decrease in dwl for the consumer in (a) must be at least as large as the (green) increase in dwl for the consumer in (b).

agent, making both agents better off. Thus, ruling out perceived price heterogeneity eliminates the possibility of an allocative inefficiency.

Proof: From lemma 2:

$$\int_{p_i^s,\theta_i,\zeta_i} [dwl(p_i^s;\theta_i,\zeta_i) + [\hat{p}^s - \bar{p}] * q(p_i^s;\theta_i,\zeta_i)]\, dF_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

$$\geq \int_{\theta_i,\zeta_i} [dwl(\hat{p}^s;\theta_i,\zeta_i) + [\hat{p}^s - \bar{p}] * q(\hat{p}^s;\theta_i,\zeta_i)]\, dF_{\theta,\zeta}\,(\theta_i,\zeta_i)$$

But note from the rationalizability of the data that:

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

Thus, we can conclude that:

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

Theorem 1 points out that, for any distribution that rationalizes the data, one can alternatively rationalize the data with a homogeneous perceived price that yields(weakly) less deadweight loss. From this, we can make two conclusions. One, we generally cannot identify deadweight loss because we could always alternatively rationalize the data with a homogeneous perceived price.[24] 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.[25]

3.2 Upper Bound on Deadweight Loss

Supposing that the tax rate τ is positive, the upper bound comes from an assumption on the limits to tax salience:

Assumption 1: There is some known value $\overline{m} \geq 0$ such that $F_{p^s}^{*}$ yields p^s with support contained entirely in $\mathcal{P} \equiv [\bar{p}, \bar{p} + m\tau]$.

This assumption, made by the econometrician, says that agents must perceive a non-negative tax τ^s no greater than $\overline{m}$ times the true tax. For instance, setting $\overline{m} = 1$ would mean assuming that agents never over-react to the tax. Imposing that $\tau^s \geq 0$ already ensures that deadweight loss is no greater than the original consumer surplus.[26] But the interval restriction implies that 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 $\partial\mathcal{P} \equiv \{\bar{p}, \bar{p} + \overline{m}\,\tau\}$.

[24] This claim holds generically, but would not hold, for instance, if there was no heterogeneity in tax salience.
[25] The existence of this lower bound is via an argument analogous to that of the upper bound.
[26] This is because deadweight loss equals its calculation as if τ^s was the true tax rate. One can show that, if τ^s has support on negative values, total deadweight loss can be substantially greater than the original total consumer surplus.

Before formally stating this result in Theorem 2, we show that one can always construct a binary distribution that rationalizes the data. Consider any $F_{p^s,\theta,\zeta}$ that rationalizes the data. If $F_{p^s,\theta,\zeta}$ puts no mass on $int(\mathcal{P}) \equiv (\bar{p}, \bar{p} + \bar{m}\,\tau)$, then the claim holds trivially. We now consider only distributions such that:

$$\lim_{m \to \bar{m}^-} F_{p^s}(\bar{p} + m\,\tau) - F_{p^s}(\bar{p}) > 0 \tag{5}$$

Pick $\tilde{p}^s \in \partial\mathcal{P}$ and a corresponding $p^b(p_i^s) \equiv \bar{p} + \mathbb{I}(p_i^s > \tilde{p}^s)\,\bar{m}\,\tau$ such that:

$$\begin{aligned}\int_{p_i^s \in int(\mathcal{P}),\theta_i} q(p^b(p_i^s);\theta_i,l)dF_{p^s,\theta}(p_i^s,\theta_i) \leq \int_{p_i^s \in int(\mathcal{P}),\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \\ \leq \int_{p_i^s \in int(\mathcal{P}),\theta_i} q(p^b(p_i^s);\theta_i,h)dF_{p^s,\theta}(p_i^s,\theta_i)\end{aligned} \tag{6}$$

In words, for any distribution that puts mass on $int(\mathcal{P})$, we pick a value $\tilde{p}^s$ that acts as a divide: people below it get assigned to a group that does not perceive the tax at all, while people 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, it is always possible to find $\lambda \in [0,1]$ such that:

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

Now, define the alternative distribution $F''_{p^s,\theta,\zeta}$ so that $F''_{p^b(p^s),\theta} = F_{p^s,\theta}$, $F''_{p^s,\theta,\zeta|p^s\in\partial\mathcal{P}} =$ $F_{p^s,\theta,\zeta|p^s\in\partial\mathcal{P}}$, and, conditional on $p^s \in int(\mathcal{P})$, $\zeta = h$ with probability λ, $\zeta = l$ with probability

$1-\lambda$, and $(p^s, \theta) \perp \zeta$. In words, we propose no change to the distribution of preference types and propose no change at all when perceived prices are at the extremes. When perceived prices are interior, the tie-breaking parameter is then independent of perceived price and preference type. Note that $F''_{p^s,\theta,\zeta}$ rationalizes the data when agents perceive subjective prices $p^b(p_i^s)$:

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

$$= \int_{p_i^s\in int(\mathcal{P}),\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)\, dF_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

$$+ \int_{p_i^s\in\partial\mathcal{P},\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)\, dF_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

$$= \int_{p_i^s\in int(\mathcal{P}),\theta_i,\zeta_i} [\lambda q(p^b(p_i^s);\theta_i,h) + (1-\lambda)\, q(p^b(p_i^s);\theta_i,l)]\, dF_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

$$+ \int_{p_i^s\in\partial\mathcal{P},\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)\, dF''_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

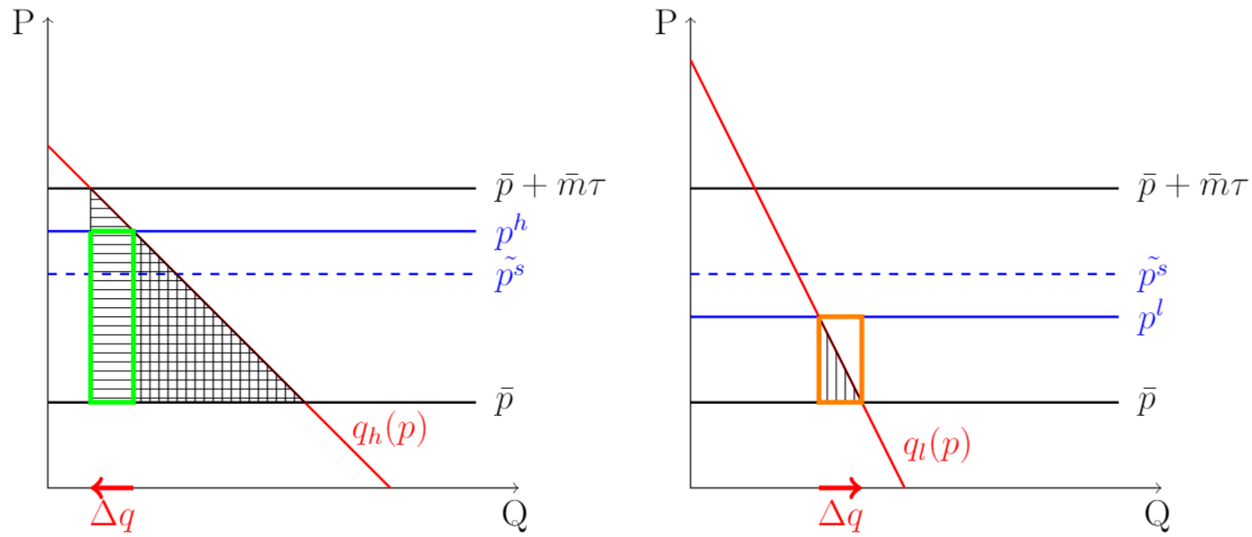
$$= \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i)\, dF''_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

Furthermore, this new distribution providers 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:

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

Figure 1.6



(a) High perceived price (b) Low perceived price

Figure 1.6: 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). Since 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.

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.

The proof follows closely the proof of theorem 1. Since we are using two separate perceived prices, $\bar{p}$ and $\bar{p} + \bar{m}\tau$, instead of just one, $\hat{p}^s$, the algebra is a bit more involved, so we relegate the proof to the appendix. Nonetheless, the idea if the proof uses lemma 2 to obtain the inequality:

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

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

$$-\int_{p_i^s,\theta_i,\zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s;\theta_i,\zeta_i)\, dF''_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

$$+\int_{p_i^s,\theta_i,\zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s;\theta_i,\zeta_i)\, dF_{p^s,\theta,\zeta}\,(p_i^s,\theta_i,\zeta_i)$$

Then, we use the CLD to show that the last two lines cancel out; because $p^b(p_i^s)$ can have two separate outputs, this is a bit more involved than in the proof of Theorem 1.

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^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. We demonstrate this point in our example in table 1. Intuitively, in a model of price misperception, people can be seemingly indifferent between several choices even when they ex-post would prefer some choices over others. This is because people might conjecture themselves different incomes, leading them to believe they can't afford their most preferred bundle. This implies that when several agents can make different choices based on their tie-breaking type ζ alone, we can increase deadweight loss by transferring some consumption from people who value it more to people who value it less, holding aggregate demand constant. Given aggregate demand and knowledge of the distribution of θ, one can always find the distribution of p^s and ζ that maximizes deadweight loss.

Theorem 3: There exists values $\Delta \in [0, \overline{m}\tau]$ and $\gamma \in [0,1]$ such that:

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

where:[27]

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

Furthermore, under assumption 1, for any $F_{p^s,\theta,\zeta}$ that rationalizes the data such that $F_\theta = F^*_\theta$:[28]

[27] Of course, if $q(\bar{p};\theta,h) = q(\bar{p}+\bar{m}\tau;\theta,l)$, then $\tilde{q}_{\Delta,\gamma}(\theta_i) = q(\bar{p};\theta_i,h)$.

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

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

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

In words, theorem 3 says that the maximal value of deadweight loss consistent with the data and knowledge of F_θ^* is given by having all agents perceive the highest possible price when doing so yields more than (and fraction γ of those for whom it's equal to) a certain value of deadweight loss per quantity reduced from the perceived price increase due to the tax. We have these same agents choose the lowest quantity consistent with their preferences and perceived budget, whereas we have the other agents choose the largest quantity consistent with their preferences and perceived budget.

We relegate this proof to the appendix, but the intuition 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.

3.3 Linear Special Case

We conclude this section with a discussion of the special case in which q is known to be a linear function of $\bar{p}$ 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. As demonstrated in section 2.3, one can express a second order approximation to deadweight loss as a function of derivatives. In the case where the choice function is linear in regressors $\bar{p}$ & τ, the second order approximation is an exact calculation of deadweight loss, and our results from the previous subsections apply.

Formally, each preference type θ_i takes the form $\theta_i = (\beta_i, \epsilon_i) \in \mathbb{R}^2$.[29] 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:

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

We are suppressing tie-breaking parameter ζ because $\theta_i, \bar{p}_i, m_i$, and τ_i always uniquely determine consumption. We have 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:

$$q_i = \alpha + \beta_i\bar{p} + \tilde{\beta}_i\tau_i + \epsilon_i$$

with corresponding deadweight loss per agent from equation (3):

$$dwl_i = \int_{\bar{p}}^{p^s} [\alpha + \beta_i p + \epsilon_i]dp - [p^s - \bar{p}] * [\alpha + \beta_i p^s + \epsilon_i] = \int_{\bar{p}}^{p^s} (p - p^s)\beta_i dp$$

$$dwl_i = \frac{1}{2}\left[\frac{p^{s2} - \bar{p}^2}{2} - [p^s - \bar{p}] * p^s\right]\beta_i = \frac{1}{2}[p^s - \bar{p}] * [(p^s + \bar{p}) - 2p^s]\beta_i = -\frac{1}{2}\tau^{s2}\beta_i$$

$$dwl_i = -\frac{1}{2}m_i^2\beta_i\tau^2$$

We assume that the joint distribution of parameters remains unaffected by the specific values of $\bar{p}$ and τ. The econometrician observes for various values of regressors:

---

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

$$\mathbb{E}[q|\bar{p},\tau] \equiv \int_{\beta_i,\tilde{\beta}_i,\epsilon_i} [\alpha + \beta_i\bar{p} + \tilde{\beta}_i\tau + \epsilon_i] dF^*_{\beta,\tilde{\beta},\epsilon}(\beta_i,\tilde{\beta}_i,\epsilon) = \alpha + \mathbb{E}[\beta]\bar{p} + \mathbb{E}[\tilde{\beta}]\tau \qquad (7)$$

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 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 $[0,\bar{m}]$.[30] The econometrician can also use the CLD as in lemma 1, 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. 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.[31] We define a measure of central tendency of tax salience:[32]

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

Then the homogeneous perceived price that rationalizes the data is $\hat{p}^s = \bar{p} + \widehat{m}\tau$. To see this, note that assuming all agents have tax salience $m_i = \widehat{m}$ yields aggregate demand as in equation (7):

[30] We assume that $\bar{m}$ is sufficiently small so that $q_i \ge 0$ with probability one, ruling out instances of negative consumption.

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

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

$$\int_{\beta_i,\epsilon_i} [\alpha + \beta_i \hat{p}^s + \epsilon_i] dF^*_{\beta,\epsilon}(\beta_i,\epsilon_i) = \alpha + \bar{p}\int_{\beta_i,\epsilon_i} \beta_i dF^*_{\beta}(\beta_i) + \widehat{m}\tau \int_{\beta_i} \beta_i dF^*_{\beta}(\beta_i) = \alpha + \hat{\beta}\bar{p} + \widehat{m}\hat{\hat{\beta}}\tau$$

$$= \alpha + \hat{\beta}\bar{p} + \hat{\hat{\beta}}\tau$$

Thus, the agent cannot rule out all agents perceiving the same price $\hat{p}^s$, and so cannot rule out $m_i = \widehat{m}\ \forall i$. For tax τ, this would yield deadweight loss, which by Theorem 1 would be a lower bound:

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

Alternatively, the econometrician cannot rule out the perceived tax τ^S having support in $\{0, \overline{m}\tau\}$. To see this, consider $\mathbb{P}(p^s = \bar{p} + \overline{m}\tau) = \frac{\widehat{m}}{\overline{m}}$ and $\mathbb{P}(p^s = \bar{p}) = 1 - \frac{\widehat{m}}{\overline{m}}$ independently of other parameters and regressors.[33] This will rationalize aggregate demand:

$$\int_{\beta_i,\epsilon_i} \left[\alpha + \beta_i \bar{p} + \frac{\widehat{m}}{\overline{m}}\ \beta_i \overline{m}\tau_i + \epsilon_i\right] dF_{\beta,\epsilon}(\beta_i,\epsilon_i) = \alpha + \mathbb{E}[\beta]\bar{p} + \widehat{m}\mathbb{E}[\beta]\tau$$

This yields deadweight loss for tax τ:

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

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

Proceeding from Theorem 2 in the previous subsection, we noted that there is a specific distribution of perceived prices on $\{\bar{p}, \bar{p} + \overline{m}\tau\}$ that maximizes deadweight loss. We describe that distribution in Theorem 3, noting that it involves assigning high or low perceived prices

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

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 $\mathbb{P}(m = \overline{m}) = \frac{\widehat{m}}{\overline{m}}$ and $\mathbb{P}(m = 0) = 1 - \frac{\widehat{m}}{\overline{m}}$ 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 $supp(m) \in \{0, \overline{m}\}$ with:

$$\mathbb{P}_F(m = \overline{m})\mathbb{E}_F\left[\tilde{\beta} | m = \overline{m}\right] = \widehat{m}\hat{\beta} = \hat{\hat{\beta}}$$

Mathematically, one can check that such a distribution rationalizes the data and yields the maximal value of deadweight loss:

$$\alpha + \mathbb{E}[\beta]\bar{p} + \mathbb{E}_F\left[\tilde{\beta}\right]\tau = \alpha + \hat{\beta}\bar{p} + \mathbb{P}_F(m = \overline{m})\mathbb{E}_F\left[\tilde{\beta} | m = \overline{m}\right]\tau = \alpha + \hat{\beta}\bar{p} + \hat{\hat{\beta}}\tau$$

$$-\frac{1}{2}\mathbb{E}_F[m^2\beta] = -\frac{1}{2}\mathbb{P}_F(m = \overline{m})\overline{m}\mathbb{E}_F\left[\tilde{\beta} | m = \overline{m}\right]\tau^2 = -\frac{1}{2}\overline{m}\hat{\hat{\beta}}\tau^2 = DWL_{high}$$

More intuitively, once one knows $\hat{\beta}$ and $\hat{\hat{\beta}}$, 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 preference doesn't matter upon attaining the observed aggregate demand.

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

$$DWL \in \left[-\frac{1}{2}\widehat{m}\hat{\hat{\beta}}\tau^2, -\frac{1}{2}\overline{m}\hat{\hat{\beta}}\tau^2\right]$$

We can conclude from this result that one cannot even identify a second order approximation of deadweight loss with aggregate data alone.[34] 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.

We showcase such an application in section B of the appendix, based on data on aggregate beer consumption from CLK (2009). We replicate the regressions they ran to estimate tax salience, but using a linear (rather than log-log) specification to match this subsection. Taking the ratio of averages across regressions of measures of aggregate demand responses to sales and excise tax variation, we estimate $\hat{m} \approx 0.27$. This estimate suggests that for $\bar{m} \geq 1$, the upper bound of deadweight loss is more than four times the lower bound.

4. Point Identification with Individual-Level Data

The previous section established that attention heterogeneity prevents point identification of deadweight loss using aggregate data, even if we already know the distribution of preferences. To facilitate identification, we require more granular data.[35] If we use (repeated) cross-sectional data, we also require additional structure on tax salience or the consumption set. Alternatively, we can use (long) panel data to identify deadweight loss without any structural assumptions.

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

[35] Technically, one could simply impose a distribution of p^s. For instance, one could assume homogeneous perceived tax-inclusive prices. We do not recommend this.

4.1 Cross-Sectional Data

We maintain the linear structure on preferences, as well as the assumption that m is constant with respect to τ, as in section 3.3. Formally:

$$q = \alpha + \beta\bar{p} + \tilde{\beta}\tau + \epsilon$$

Here α is a constant whose identification we generally assume.[36] The econometrician observes the distribution of q conditional on $(\bar{p}, \tau)$ for all values in $supp(\bar{p}, \tau)$. The data comes from an underlying data-generating process, which we assume yields a well-defined and finite value for deadweight loss. The data identifies expected deadweight loss from a (non-zero) tax τ if any underlying distributions of $(\beta, \tilde{\beta}, \epsilon)$ across the population consistent with the observed distribution of $(\bar{p}, \tau)$ and conditional distributions of q yield the same value of $\mathbb{E}\left[\frac{\tilde{\beta}^2}{\beta}\right]$. One approach is to identify the joint distribution of $(\beta, \tilde{\beta})$ across individuals, then integrate to obtain the desired expected value. To that end, we can use the following lemma from Masten (2017):

Lemma 3: If $supp(\bar{p}, \tau)$ contains an open ball in $\mathbb{R}^2$, then the joint distribution of $(\beta, \tilde{\beta}, \epsilon)$ is identified if and only if:

1. The joint distribution of $(\beta, \tilde{\beta}, \epsilon)$ is determined by its moments.
2. All absolute moments of $(\beta, \tilde{\beta}, \epsilon)$ are finite.

Proof: Apply lemma 2 from Masten (2017). ■

[36] It is identified if the econometrician observes $\bar{p} = \tau = 0$ or non-collinear variation in $\bar{p}$ and τ.

It follows as an immediate corollary that the conditions of the lemma also identify deadweight loss, since one can integrate out the marginal distribution of $(\beta, \tilde{\beta})$ and then calculate $\mathbb{E}\left[\frac{\tilde{\beta}^2}{\beta}\right]$. The intuition is that one can find the moments of the random coefficients by running increasingly higher-order regressions of the form:

$$\mathbb{E}[q_i^n | \bar{p}, \tau] = \cdots + \mathbb{E}[\beta^n]\bar{p}^n + \cdots + \mathbb{E}[\tilde{\beta}^n]\tau^n + \cdots$$

The enumerated conditions are satisfied, for instance, when $(\beta, \tilde{\beta})$ has finite support. However, one might want to be able to identify deadweight loss without making assumptions on the distribution of $(\beta, \tilde{\beta})$, beyond assuming that $\mathbb{E}\left[\frac{\tilde{\beta}^2}{\beta}\right]$ is well-defined and finite. The following theorem does not impose such assumptions, but does require unbounded support in regressors.

Theorem 4: Suppose that for every pair $(\lambda_1, \lambda_2) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R}$, there is a sequence $(\bar{p}_k, \tau_k)_{k=1}^{\infty}$ contained within the support of $(\bar{p}, \tau)$ such that:

1. $\lim_{k \to \infty} \|(\bar{p}_k, \tau_k)\| = \infty$
2. $\lim_{k \to \infty} \frac{\tau_k}{\bar{p}_k} = \frac{\lambda_2}{\lambda_1}$

In addition, suppose that there is another sequence $(\bar{p}_k, \tau_k)_{k=1}^{\infty}$ contained within the support of $(\bar{p}, \tau)$ such that:

1. $\lim_{k \to \infty} |\bar{p}_k| < \infty$
2. $\lim_{k \to \infty} \tau_k = \infty$

Then $\mathbb{E}[dwl_i]$ is identified.

We make note of a special case in which these results apply. Consider a binary choice problem, so that $X^T = \{0,1\}$. Each agent i has quasi-linear utility u_{i1} fro consuming the good, so that for tax-inclusive price p:

$$u_{i1} = \epsilon_i - p$$

This is an expression of the agent's preference for the taxed good scaled so that an additional dollar yields one unit of utility. We do not assume that $\mathbb{E}[\epsilon] = 0$. However, we normalize utility in the absence of the taxed good to zero:

$$u_{i0} = 0$$

For a price increase from $\bar{p}$ to p^s, the change in the expenditure function is:

$$e(p^s) - e(\bar{p}) = \begin{cases} 0, & \epsilon_i - \bar{p} \leq 0 \\ \epsilon_i - \bar{p}, & p^s \geq \epsilon_i > \bar{p} \\ p^s - \bar{p}, & \epsilon_i - p^s > 0 \end{cases}$$

Each agent i has a tax salience m_i independent of the tax rate. The agent perceives utility from buying the taxed good:

$$u_{i1}^s = \epsilon_i - \bar{p} - m_i \tau \tag{8}$$

They purchase the good if $u_{i1}^s > 0$. This yields deadweight loss for that agent of:

$$dwl_i = e(p^s) - e(\bar{p}) - [p^s - \bar{p}]\mathbb{I}(\alpha - p^s + \epsilon_i > 0) = \begin{cases} 0, & \epsilon_i - \bar{p} \leq 0 \\ \epsilon_i - \bar{p}, & p^s \geq \epsilon_i > \bar{p}\text{s} \\ 0, & \epsilon_i - p^s > 0 \end{cases}$$

Let $F_{m.\epsilon}^*$ denote the true distribution of (m, ϵ). For simplicity, we assume that $\mathbb{P}(u_1^s = 0)$ and $\mathbb{P}(u_1 = 0)$ are always zero. We assume the econometrician observes only aggregate data, so that the only values observed are triplets $\left(\bar{p}, \tau, \mathbb{P}(u_1^s > 0 | \bar{p}, \tau)\right)$. The challenge is to infer deadweight loss, which takes the form:

$$DWL = \int_{m_i, \epsilon_i} \mathbb{I}(u_1 \geq 0 > u_1^s)[\epsilon_i - \bar{p}] dF_{m,\epsilon}^*(m_i, \epsilon_i) \tag{9}$$

For (2017) shows that, if $\bar{p}$ has full support, then one can use aggregate demand to infer the CDF of $\epsilon - m\tau$, allowing one to apply Masten (2017). Thus, we can identify deadweight loss even with aggregate data if we are willing to assume that tax salience does not depend on the tax rate and the choice set is binary.

One might hope that the assumption that $m \perp (\bar{p}, \tau)$ is not required for identification, as it is generally a rather strong assumption. Unfortunately, one cannot do away entirely with this restriction. For instance, consider a population in which $\mathbb{P}(\epsilon = 2) = \mathbb{P}(\epsilon = 3) = 0.5$ and $\mathbb{P}(\zeta = l) = 1$. Consider rationalizing the data in two ways. In the first case, $m = 0.5$ with probability one. This yields aggregate demand:

$$D = \begin{cases} 1, & \bar{p} + 0.5\tau \leq 2 \\ 0.5, & 3 \geq \bar{p} + 0.5\tau > \bar{2} \\ 0, & \bar{p} + 0.5\tau > 3 \end{cases}$$

Aggregate deadweight loss takes the form:

$$DWL = \begin{cases} 0, & \bar{p} + 0.5\tau \leq 2 \\ 0.5, & 3 \geq \bar{p} + 0.5\tau > \bar{2} \\ 1.5, & \bar{p} + 0.5\tau > 3 \end{cases}$$

In the second case, suppose for a moment that $\bar{p} \in (0,1)$. Then agents with $\epsilon = 2$ have tax salience $m_2(\bar{p}) = \frac{2-\bar{p}}{6-2\bar{p}}$, while agents with $\epsilon = 3$ have tax salience $m_3(\bar{p}) = \frac{3-\bar{p}}{4-2\bar{p}}$. If $\bar{p} \notin (0,1)$, then set $m_2(\bar{p}) = m_3(\bar{p}) = 0.5$. One can check that this yields the same aggregate demand, yet aggregate deadweight loss with $\bar{p} \in (0,1)$ now takes the form:

$$DWL = \begin{cases} 0, & \bar{p} + 0.5\tau \leq 2 \\ 0.75, & 3 \geq \bar{p} + 0.5\tau > \bar{2} \\ 1.5, & \bar{p} + 0.5\tau > 3 \end{cases}$$

4.2 Panel Data

Now suppose that one can follow specific individual for a long period of time. For every individual i, the econometrician observes triplets $(\bar{p}_i, \tau_i, q_i)$ with full support for $\bar{p}$ when $\tau = 0$. Then the econometrician can identify the demand function $q_i(p; 0)$. Assuming there are no income effects, deadweight loss for an individual i with perceived price p^s takes the form:

$$dwl_i = \int_{\bar{p}}^{p^s} q(p; 0) dp - \tau d(p^s)$$

If one can observe tax τ, then one can determine deadweight loss using the inferred demand function via:[37]

$$p^s = d^{-1}\big(q_i(\bar{p}, \tau)\big)$$

For instance, with binary choice data, demand for agent i is $\mathbb{I}(\alpha - p_i^s + \epsilon_i)$. This yields deadweight loss for agent i of:[38]

$$dwl_i = \mathbb{I}(p_i^s > \alpha + \epsilon_i > \bar{p})\ [\alpha - \bar{p}\ \epsilon_i]$$

This yields aggregate deadweight loss:

$$DWL = \int_{p_i^s, \epsilon_i} \mathbb{I}(p_i^s > \alpha + \epsilon_i > \bar{p})\ [\alpha - \bar{p} + \epsilon_i]\ dF^*_{p^s,\epsilon}(p_i^s, \epsilon_i)$$

5. Empirical Calculation

We apply our theoretical results to experimental data from Taubinsky and Rees-Jones (forthcoming). We use a subset of their data in which subjects are randomly divided into two groups of about 1,000 test subjects. For each of twenty goods, one group reports their maximal willingness to pay for that good, while the other group reports the maximal sticker price at

[37] If demand is a correspondence, then one can find p^s as the price for which $q_i(\bar{p}, \tau)$ is an element of the demand correspondence at p^s. By lemma 1, this value is unique.

[38] We assume that demand is well-defined with probability one.

which they would be willing to purchase the good subject to the sales tax. The test subjects' city of residence determines the size of the sales tax.[39]

Data from the experiment allow us to infer aggregate demand for people in the sample. We compare the standard tax arm to the no-tax arm for a single good.[40] The tax arm allows us to consider what aggregate demand is at a tax of τ, equal to the sales tax in vigor in each subject's city of residence. Then we turn to the no-tax arm, which allows us to observe people's willingness to pay in the absence of taxation. We interpret this as the subjects' *true* willingness to pay, meaning that if their declared willingness to pay is ϵ, then they will buy the good if the price they *perceive* is lower than ϵ. In turn, this allows us to find the perceived prices $\hat{p}^s$ from equation (4) and $\tilde{p}^s$ from equation (6) that allow us to apply Theorems 1 and 3.

Specifically, we compute aggregate demand under the tax, d^*, by finding the fraction of people in the tax arm who declare a higher willingness to pay than the true sticker price for the object (taken from Amazon). We then proceed to find the homogeneous perceived price that would equalize the demand for the no-tax arm to d^*. Letting ϵ_i indicate the willingness to pay for subject i in the no-tax arm, we then find $(\hat{p}, \hat{\lambda})$ that solves:

$$d^* = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(\epsilon_i > \hat{p}) + \hat{\lambda}\mathbb{I}(\epsilon_i = \hat{p})$$

Where $\hat{\lambda}$ is introduced to deal with edgewise cases and satisfy the equation with equality. One could interpret $\hat{\lambda}$ as the probability that $\zeta = h$ in the no-tax sample, or the fraction of

[39] We describe only the data that we use, which is not all the data from the experiment. For instance, Taubinsky and Rees-Jones (forthcoming) have a second module in which they ask both groups their willingness to pay in the absence of any sales taxes.

[40] The good we consider is RainStoppers 68-inch oversize windproof golf umbrella.

indifferent people who end up buying. This, in turn, allows us to find the lower bound of deadweight loss as in equation (9):[41]

$$DWL^{l} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathbb{I}(\bar{p} > \epsilon_i < \hat{p}) + \left(1-\hat{\lambda}\right)\mathbb{I}(\epsilon_i = \hat{p})\right] * [\epsilon_i - \bar{p}]$$

Finding the upper bound of deadweight loss follows a similar procedure. The problem is then to find $\left(\tilde{p}, \tilde{\lambda}\right)$ such that:

$$d^{*} = \frac{1}{N}\sum_{i=1}^{N}(\bar{p} > \epsilon_i < \tilde{p}) + \tilde{\lambda}\,\mathbb{I}(\epsilon_i = \tilde{p})$$

Demand is rationalized by having all agents with ϵ_i less than $\tilde{p}$ perceive price $\bar{p}$ and all agents with $\epsilon_i > \tilde{p}$ perceive a price equal t the supremal value of the support of ϵ. The upper bound for deadweight loss is then what we would obtain if all the subjects who value the good most perceived a price so high that it would dissuade them from buying the good:

$$DWL^{u} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathbb{I}(\epsilon_i > \tilde{p}) + \left(1-\tilde{\lambda}\right)\mathbb{I}(\epsilon_i = \tilde{p})\right] * [\epsilon_i - \bar{p}]$$

The calculations are very similar when we assume that $\bar{m}=1$, except in that case anyone with $\epsilon_i > \tau$ is assumed to always buy, thus resulting in a smaller upper bound. We also compute the implied value of deadweight loss one would obtain if preferences and salience are independent. In the case where the econometrician does not impose any assumptions on what $\bar{m}$ might be, so that anyone who perceives the high price will not buy the good, the problem reduces to picking α so that:

[41] We are calculating deadweight loss as if there were no other taxes in the economy.

Table 2

	Full salience	Naive	Lower Bound	UB1	UB2	UB3	UB4
Dollar value	0.187	0.0266	0.0038	0.0068	0.0441	0.0100	0.2097
Percent of revenue	10.55%	15.01%	2.14%	3.84%	24.89%	5.64%	118.34%

Table 2: Lower and upper bounds of deadweight loss, as inferred from the data in Taubinsky and Rees-Jones (forthcoming). We consider one lower bound and four different upper bounds. UB1 is the upper bound one would obtain assuming that salience is independent of preferences and that maximal salience is 1 (that is, that people do not over-react to the tax). UB2 is the upper bound one would obtain only assuming that salience is independent of preferences. UB3 is the upper bound one would obtain assuming only that maximal salience is 1. UB4 is the highest possible upper bound. We also report what deadweight loss would be if we assumed that people in the non-taxed sample accounted for taxes perfectly, and the naive calculation the econometrician would obtain if they assumed the observed reaction to taxes in the taxed sample was induced by full salience. For comparison, the implied average tax revenue from demand under the tax is 0.1772, and the implied total pre-tax consumer surplus is 0.4866.

$$d^* = \frac{1}{N}\sum_{i=1}^{N} \alpha \mathbb{I}(\epsilon_i \geq \bar{p})$$

And the resulting upper bound on deadweight loss will be:

$$DWL^u_{m\perp\epsilon} = (1-\alpha)\frac{1}{N}\sum_{i=1}^{N} \mathbb{I}(\epsilon \geq \bar{p})\,[\epsilon_i - \bar{p}]$$

We summarize our results in table 2.[42] As we can see, the lower and upper bounds are quite far apart, implying that the exact co-distribution of preferences and perceived prices can have widely different implications for welfare. One could find an analogue for $\hat{m}$ by dividing the lower bound by the upper bound in the case of $\bar m = 1$. This would yield $\hat{m} = 0.38$, similar to $\hat{m} \approx 0.27$ as in section 3.3.

[42] One limitation of our result is that we treat the sample as if it was the population in question. We conjecture that one could obtain standard errors for estimates by using asymptotic normality from the Generalized Method of Moments estimator. We intend to prove this claim and include standard errors in future versions of this paper.

Our empirical results suggest two takeaways. One, we require strong assumptions to get an upper bound for deadweight loss close to the lower bound. To rule out deadweight loss multiple times greater than the deadweight loss with a homogeneous perceived price, the econometrician must assume both that tax salience is bounded between zero and one, and is also distributed independently of the willingness to pay. Two, comparing UB1 to UB2 and UB3 suggests that that the potential dependent statistical relationship between tax salience and willingness to pay does not facilitate deadweight loss nearly as much as the possibility of agents perceiving a tax rate greater than the true tax rate. This motivates us to investigate how the upper bound of deadweight loss varies for different choices for maximal salience $\overline{m}$, reported in figure 1.7.

The high variation in empirical estimates suggests that precisely inferring deadweight loss from aggregate data would require strong assumptions. On the other hand, for most reasonable values of $\overline{m}$, deadweight loss seems relatively small. In turn, this suggests that attempting to raise the same revenue with a less distortive tax scheme might not be worth it, if it incurs high enough administrative costs.[43]

6. Conclusion

In this paper, we considered identification of deadweight loss with non-salient taxes, as in the Gabaix (2014) model. We justify considering price misperception by demonstrating that a general model of consumer behavior with weak convexity and continuity assumptions on preferences is observationally equivalent to the Gabaix (2014) model.

[43] Formally, if the administrative cost of implementing a more efficient tax that collected the same revenue was greater than around 32 cents per consumer in the market, one could conclude with high confidence that the change in tax regime could not yield a Pareto improvement with transfers.

Figure 1.7

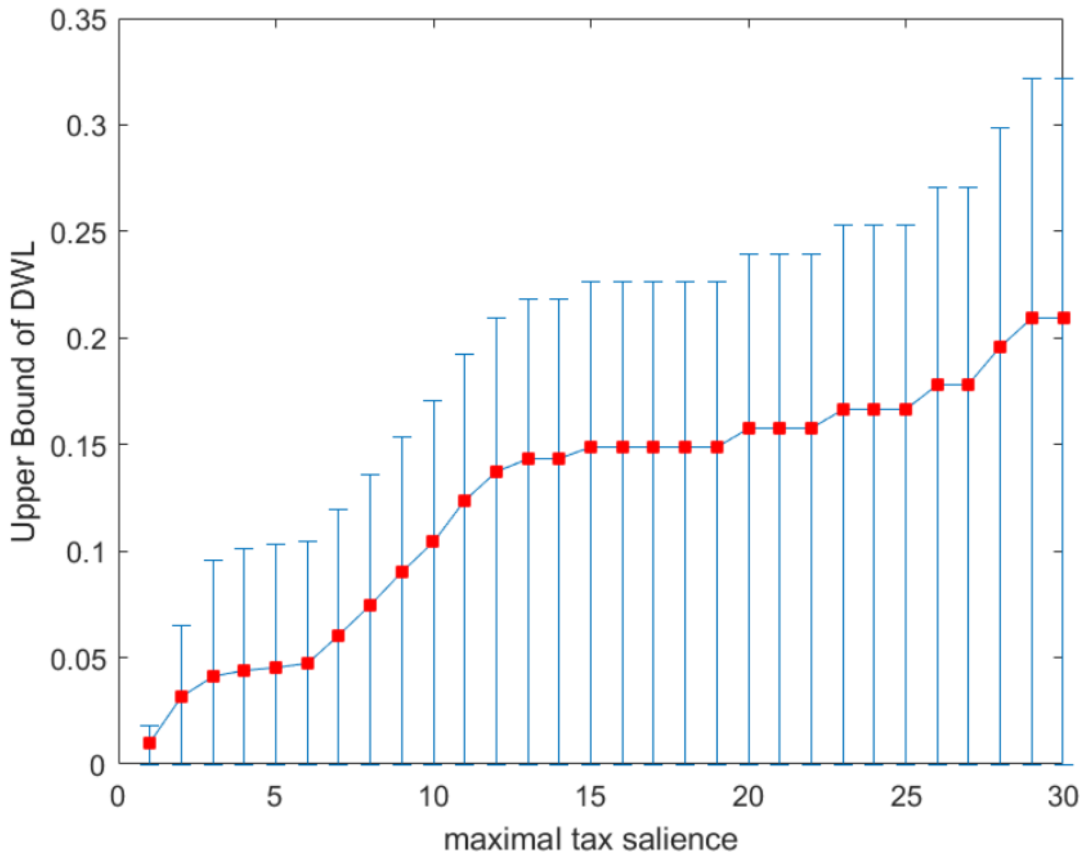


Figure 1.7: Upper bound depending on the econometrician's choice of $\overline{m}$. Error bars are based on the 5th and 95th percentiles of 10,000 bootstrap replications.

We first consider identification of deadweight loss using aggregate data. We show that deadweight loss cannot be point-identified with aggregate demand data. Nonetheless, we provide bounds for deadweight loss consistent with aggregate demand and the distribution of preference parameters across individuals. The lower bound holds for any distribution; the upper bound relies on the assumption that tax salience has support contained in a known non-negative interval. If the econometrician could not infer the distribution of preference parameters or did not know a non-negative interval that contained the support of tax salience, then the interval of possible values of deadweight loss may be even larger.

We provide context for these theoretical results with empirical findings using experimental data from Taubinsky and Rees-Jones (forthcoming). We calculate the upper and lower bounds of deadweight loss due to sales taxes in the United States on a good with a binary choice set. We find an upper bound approximately fifty-five times larger than the lower bound. Interestingly, this upper bound is substantially larger than the deadweight loss if all agents perfectly accounted for sales taxes. In other words, deadweight loss with imperfectly perceived taxes can be greater than in the standard model of consumer behavior. This result arises from the possibility that some agents that value the good a lot perceive a prohibitively high price, whereas other agents who value the good at barely more than the sticker price still purchase the good because they fail to notice the tax at all. Thus, the behavioral model does not rule out deadweight loss greater than in the standard model even though aggregate consumption is less distorted in the behavioral model. The allocative inefficiency swamps the distortion in aggregate consumption. Indeed, even when no agents respond more to sales taxes than sticker price changes, we find for the binary and continuous choice settings we study that deadweight loss can be roughly as large as two & a half and four times as large as when assuming homogeneous tax salience, respectively.

With individual-level cross-sectional data, identification arises from assuming choice functions that are linear in sticker prices and taxes, and that tax salience is independent of sticker prices and taxes. Under these assumptions, one can identify the joint distribution of parameters of demand responsiveness to sticker prices and taxes, and so compute aggregate deadweight loss. With long panel data, we can identify deadweight loss without these assumptions. Identification then arises from direct calculation of the lost surplus from agents facing the tax compared to when there was no tax.

In future work, we will explore the performance of estimators of deadweight loss using individual-level cross-sectional data. The assumptions on preferences and tax salience facilitate identification, but they leave open two questions. One, our identification argument in Theorem 4 does not inform us about rates of convergence. Two, the structural assumptions may not hold in practice, and we would like to have some sense of the magnitude of the bias that these assumptions introduce.

We also hope to determine to what degree one can identify deadweight loss in the intermediate case of individual-level cross-sectional data in which tax salience depends arbitrarily on the tax rate and preferences, but does not depend on the sticker price. This assumption of independence of the tax salience from the sticker price seems like the most reasonable non-trivial independence assumption one could impose on tax salience.

Finally, future work should inform what reasonable ex-ante restrictions on tax salience one can impose when observing aggregate data. Our empirical results show widespread uncertainty as to what deadweight loss might be with aggregate data in the absence of restrictions on tax salience. In fact, we still assume that tax salience is non-negative with probability one. Relaxing this assumption could yield an even greater upper bound for deadweight loss.

References

Beran, Rudolf, and Peter Hall. "Estimating Coefficient Distributions in Random Coefficient Regressions." *The Annals of Statistics* (1992): 1970-1984.

Beran, Rudolf, Andrey Feuerverger, and Peter Hall. "On Nonparametric Estimation of Intercept and Slope Distributions in Random Coefficient Regression." *The Annals of Statistics* 24.6 (1996): 2569-2592.

Berry, Steven, James Levinsohn, and Ariel Pakes. "Automobile Prices in Market Equilibrium." *Econometrica: Journal of the Econometric Society* (1995): 841-890.

Berry, Steven, James Levinsohn, and Ariel Pakes. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The new car market." *Journal of Political Economy* 112.1 (2004): 68-105.

Chetty, Raj, Adam Looney, and Kory Kroft. "Salience and Taxation." *NBER Working Paper W* 13330 (2007).

Chetty, Raj, Adam Looney, and Kory Kroft. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99.4 (2009): 1145-77.

Gabaix, Xavier, and Emmanuel Farhi. "Optimal Taxation with Behavioral Agents." *2017 Meeting Papers*. No. 1634. Society for Economic Dynamics, 2017.

Finkelstein, Amy. "E-ztax: Tax Salience and Tax Rates." *The Quarterly Journal of Economics* 124.3 (2009): 969-1010.

Fox, Jeremy T. *A Note on Nonparametric Identification of Distributions of Random Coefficients in Multinomial Choice Models*. No. w23621. National Bureau of Economic Research, 2017.

Gabaix, Xavier. "A Sparsity-Based Model of Bounded Rationality." *The Quarterly Journal of Economics* 129.4 (2014): 1661-1710.

Gallagher, Kelly Sims, and Erich Muehlegger. "Giving Green to Get Green? Incentives and Consumer Adoption of Hybrid Vehicle Technology." *Journal of Environmental Economics and Management* 61.1 (2011): 1-15.

Goldin, Jacob. "Optimal Tax Salience." *Journal of Public Economics* 131 (2015): 115-123.

Goldin, Jacob, and Tatiana Homonoff. "Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity." *American Economic Journal: Economic Policy* 5.1 (2013): 302-36.

Hoderlein, Stefan, Jussi Klemelä, and Enno Mammen. "Analyzing the Random Coefficient Model Nonparametrically." *Econometric Theory* 26.3 (2010): 804-837.

Rosen, Harvey S. "Tax Illusion and the Labor Supply of Married Women." *The Review of Economics and Statistics* (1976): 167-172.

Taubinsky, Dmitry, and Alex Rees-Jones. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies* (forthcoming).

Appendix for Chapter 1

A. Additional Proofs and Theoretical Results

A.1 Additional Proofs and Results from Section 2

We first demonstrate a generalization of the Gabaix representation, in which multiple goods may be taxed. We consider a general setting with N goods, consumption set $X \equiv X^T \times X^{NT} \subseteq \mathbb{R}_+^N$, with consumption vector $\boldsymbol{q} = (\boldsymbol{q}^T, \boldsymbol{q}^{\boldsymbol{NT}}) \in X$. 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 $\succcurlyeq$ 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:

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

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 $\lambda_1, \ldots, \lambda_n \geq 0$ and $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n \in \mathcal{A}$, if $\sum_{k=1}^{n} \lambda_k = 1$, then:

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

We provide this description of CPC preferences to facilitate intuition, but we demonstrate our main result using an equivalent, yet more geometric, description of CPC.

Lemma 4: Preferences $\succcurlyeq$ 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 $\mathcal{O} \cap X = \mathcal{A}$.[44]

Proof: For one direction, the convex hull of the union of open sets $\epsilon(\boldsymbol{q}')$ 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 $\boldsymbol{q}'' \succ \boldsymbol{q}'$. ■

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. Let $\overline{\boldsymbol{p}} = (\overline{\boldsymbol{p}}^T, \overline{\boldsymbol{p}}^{NT})$ denote the vector of sticker prices.

[44] Note that $\mathcal{O}$ is open in $\mathbb{R}^N$, so that $\mathcal{O} \cap X$ is open in X.

Let $\boldsymbol{\tau}$ denote the vector of (per-unit) taxes of dimension N. The consumption vector $\boldsymbol{q}(\overline{\boldsymbol{p}}, \boldsymbol{\tau}) = \left(\boldsymbol{q}^T(\overline{\boldsymbol{p}}, \boldsymbol{\tau}), \boldsymbol{q}^{NT}(\overline{\boldsymbol{p}}, \boldsymbol{\tau})\right) \in X$ satisfies the following properties:

$$\overline{\boldsymbol{p}} * \boldsymbol{q}^{NT}(\overline{\boldsymbol{p}}, \boldsymbol{\tau}) \leq W - \boldsymbol{p}^T * \boldsymbol{q}^T$$

$$(\boldsymbol{q}^T, \boldsymbol{q}^{NT}) \succ (\boldsymbol{q}^T, \widehat{\boldsymbol{q}}^{NT}) \; \forall \widehat{\boldsymbol{q}}^{NT} \in X^{NT}: \overline{\boldsymbol{p}}^{NT} * \widehat{\boldsymbol{q}}^{NT} \leq W - \boldsymbol{p}^T * \boldsymbol{q}^T$$

$$\boldsymbol{q}(\overline{\boldsymbol{p}}, \boldsymbol{0}) \in \underset{\widehat{\boldsymbol{q}} \in X: \boldsymbol{p} * \widehat{\boldsymbol{q}} \leq W}{\operatorname{argmax}} \succcurlyeq$$

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.:

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

The claim is that for any $\overline{\boldsymbol{p}}$ and $\boldsymbol{\tau}$ in this domain, there is a Gabaix representation for $\boldsymbol{q}(\overline{\boldsymbol{p}}, \boldsymbol{\tau})$.

Proof of Generalization of Proposition 1: Define $\boldsymbol{q} = (\boldsymbol{q}^T, \boldsymbol{q}^{NT}) = \boldsymbol{q}(\overline{\boldsymbol{p}}, \boldsymbol{\tau})$, and:

$$\mathcal{A}^e \equiv \left\{(\boldsymbol{q}^{T\prime}, e^{NT\prime}) \mid \boldsymbol{q}^{T\prime} \in X^T, \exists \boldsymbol{q}^{NT\prime} \in X^{NT}: \overline{\boldsymbol{p}}^{NT} * \boldsymbol{q}^{NT\prime} = e^{NT\prime}, \left(\boldsymbol{q}^{T\prime}, \boldsymbol{q}^{NT\prime}\right) \in \mathcal{O}\right\}$$

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

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

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

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

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

Pseudo-convexity of preferences implies that:

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

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

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

$$\overline{\boldsymbol{p}}^{NT} * \sum_{k=1}^{n} \lambda_k \boldsymbol{q}_k^{NT} = \sum_{k=1}^{n} e_k^{NT} = \overline{\boldsymbol{p}}^{NT} * \boldsymbol{q}^{NT}$$

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

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:

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

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

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

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 $\overline{\boldsymbol{p}}^{NT} > \mathbf{0}$ 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 $\forall \boldsymbol{q}' \in X$:

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

Therefore, the Gabaix 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:

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

This makes it clear that we have, in fact, generalized Proposition 1. 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 nontaxed 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 on-taxed goods chosen from $\{0,1\}$. The sticker price vector is $\overline{\boldsymbol{p}} = (1,1,1)$. The consumer has 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). Consider 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.

Details from Table 1: Let $\bar{p} = \tau = p^{NT} = \overline{m} = 1$, $X^T = \{0,1\}$, and $u(q^T, q^{NT}; \theta_i) = \theta_i * q^T + q^{NT}$. In words, the marginal cost of the taxed good is zero, the tax-inclusive prices are one, salience is between zero & one, the choice of the taxed good is either zero or one, and the taxed & non-taxed goods are symmetric perfect substitutes. Suppose:

$$F_\theta(\theta_i) = 0.5\mathbb{I}(\theta_i \geq 1) + 0.5\mathbb{I}(\theta_i \geq 0)$$

This says that $\theta \in \{0,1\}$ with equal probability. We consider two distinct methods of distributing the perceived prices on $\{0,1\}$ dependent on preference type to rationalize an aggregate demand of 0.5. First, suppose that $\theta_i = 1 \Rightarrow p_i^s = 1, \zeta_i = h$, and $\theta_i = 2 \Rightarrow p_i^s = 2$, $\zeta_i = l$. Then:

$$DWL = 0.5[e(1;1) - e(1;1) - 0] + 0.5[e(2;2) - e(1;2) - 0] = 0.5$$

Second, suppose that $\theta_i = 1 \Rightarrow p_i^s = 1, \zeta_i = l$ and $\theta_i = 2 \Rightarrow p_i^s = 2, \zeta_i = h$. Then:

$$DWL = 0.5[e(1;1) - e(1;1)] + 0.5[e(2;2) - e(1;2) - 1] = 0$$

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:

$$p_i^s(\bar{p},0) - \bar{p}\ \forall\bar{p}\ \forall i$$

We implicitly define the pre-tax sticker price $\bar{p}^{old}$ by:[45]

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

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

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

Define deadweight loss by:[46]

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

where

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

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

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

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

Proof:

[45] $\mathcal{v}_i \equiv u_i\big(\boldsymbol{d}_i(\boldsymbol{p},W_i)\big)\ \forall i$

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

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

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:[47]

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

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

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

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

$$- \sum_i \left[\frac{\partial p_i^s}{\partial \tau} h_i + [p_i^s(\bar{p}^{new},\tau) - \bar{p}^{new}]\frac{\partial h_i}{\partial p}\left[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}\right]\right]$$

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

$$- \sum_i [p_i^s(\bar{p}^{new},\tau) - \bar{p}^{new}]\frac{\partial h_i}{\partial p}\left[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}\right]$$

$$= -\sum_i [p_i^s(\bar{p}^{new},\tau) - \bar{p}^{new}]\frac{\partial h_i}{\partial p}\left[\frac{\partial \bar{p}^{new}}{\partial \tau} + \frac{\partial p_i^s}{\partial \tau}\right]$$

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

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

[47] 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:

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

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 $\frac{\partial h_i}{\partial p}$ are also continuous:

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

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

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:

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

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

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

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

A.2 Additional Proofs and Results from Section 3

Proof of Lemma 1: Note that there must be values q^{NT} and $q^{NT\prime}$ such that:

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

From local non-satiation:

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

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

Rearranging yields:

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

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

Proof of Proposition 3: From lemma 1 and prices being bounded away from zero, we can always find a value of $\hat{p}^s$ such that:

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

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

$$\lambda \int_{\theta_i} q(\hat{p}^s;\theta_i,h)dF_\theta(\theta_i) + (1-\lambda)\int_{\theta_i} q(\hat{p}^s;\theta_i,l)dF_\theta(\theta_i)$$

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

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

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

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

Proof of Theorem 2: From lemma 2 and rationalizability of the data:

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

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

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

$$\geq \int_{p_i^s,\theta_i,\zeta_i} [dwl(p_i^s;\theta_i) + p_i^s\, q(p_i^s;\theta_i,\zeta_i)]\, dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

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

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

Rearranging yields:

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

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

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

$$+ \int_{p_i^s,\theta_i,\zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s;\theta_i,\zeta_i) dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i)$$

We can show from lemma 1 and $p_i^s \in [\bar{p}, \bar{p} + \bar{m}\tau] = \mathcal{P}\ \forall i$ that the last two terms total to a a non-negative value. Formally, for any $p_i^s \in (\bar{p}, \bar{p} + \bar{m}\,\tau), \theta_i, \zeta_i, \zeta_i'$:

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

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

Either way:

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

Thus:

$$\int_{p_i^s, \theta_i, \zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s; \theta_i, \zeta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$= \int_{p_i^s \in int(\mathcal{P}), \theta_i, \zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s; \theta_i, \zeta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$+ \int_{p_i^s \in \partial\mathcal{P}, \theta_i, \zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s; \theta_i, \zeta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$\geq \int_{p_i^s \in int(\mathcal{P}), \theta_i, \zeta_i} [p_i^s - \tilde{p}^s]\, q(p^b(p_i^s); \theta_i, \zeta_i) dF''_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$+ \int_{p_i^s \in \partial\mathcal{P}, \theta_i, \zeta_i} [p_i^s - \tilde{p}^s]\, q(p_i^s; \theta_i, \zeta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$= \int_{p_i^s, \theta_i} [p_i^s - \tilde{p}^s]\, q(p^b(p_i^s); \theta_i, \zeta_i) dF''_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i)$$

$$\int_{p_i^s, \theta_i} dwl(p^b(p_i^s); \theta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i) \geq \int_{p_i^s, \theta_i, \zeta_i} dwl(p_i^s; \theta_i) dF_{p^s,\theta,\zeta}(p_i^s, \theta_i, \zeta_i) \qquad \blacksquare$$

We now proceed to proving Theorem 3, i.e. that deadweight loss is maximized given the available data and distribution F_θ^* by having a cutoff value Δ such that the ratio of $dwl(\bar{p} + \bar{m}\tau; \theta_i, h)$ to $q(\bar{p} + \bar{m}\tau; \theta_i, l) - q(\bar{p} + \bar{m}\tau; \theta_i, h)$ is greater (less than) Δ for those we assign a perceived price of $\bar{p} + \bar{m}\tau$ (respectively $\bar{p}$). However, we must also consider the choice of conditional distribution of ζ. Toward that end, we note that deadweight loss is bounded by the product of the reduction in demand and $\bar{m}\,\tau$.

Lemma 5: 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 2:

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

$$dwl(p_i^s;\theta_i,\zeta_i) \leq [q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)] * [p_i^s - \bar{p}] \leq [q(\bar{p};\theta_i,\zeta_i) - q(p_i^s;\theta_i,\zeta_i)]\, \bar{m}\tau \quad \blacksquare$$

Proof of Theorem 3: The outline of the proof is as follows. First, we use lemma 5 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:

$$\begin{aligned} \rho \int_{\theta_i} q(\bar{p} + \bar{m}\tau; \theta_i, l)\, dF_{\theta|p^s \neq \bar{p}}(\theta_i) + (1-\rho) \int_{\theta_i} q(\bar{p}; \theta_i, h) dF_{\theta|p^s = \bar{p}}(\theta_i) \\ = \int_{p_i^s,\theta_i,\zeta_i} q(p_i^s;\theta_i,\zeta_i) dF_{p^s,\theta,\zeta}(p_i^s,\theta_i,\zeta_i) \end{aligned} \tag{10}$$

Such a value of ρ must exist by the Intermediate Value Theorem, since by the definitions of l & h and the CLD as expressed in lemma 1:

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

$$\leq \int_{\theta_i} q(\bar{p}; \theta_i, h) dF_{\theta|p^s = \bar{p}}(\theta_i)$$

In words, we are constructing an alternative distribution that rationalizes aggregate demand such that $p^s = \bar{p} + \bar{m}\tau$ & $\zeta = l$ with probability ρ, and otherwise $p^s = \bar{p}$ & $\zeta = h$. 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:

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

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

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:

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

$$= \rho \int_{\theta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,l) dF_{\theta|p^s\neq\bar{p}}(\theta_i)$$

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

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

$$+ F_{p^s}(\bar{p}) \int_{\theta_i,\zeta_i} dwl(\bar{p};\theta_i,\zeta_i) dF_{\theta,\zeta|p^s=\bar{p}}(\theta_i,\zeta_i)$$

$$= \int_{\theta_i,\zeta_i} dwl(\bar{p}+\bar{m}\tau;\theta_i,\zeta_i) dF_{\theta,\zeta}(\theta_i)$$

The inequality follows from $\rho \geq 1 - F_{p^s}(\bar{p})$ by assumption. 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 5:

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

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

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

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

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

$$+ F_{p^s}(\bar{p}) \int_{\theta_i,\zeta_i} q(\bar{p}; \theta_i, \zeta_i) dF_{\theta,\zeta|p^s = \bar{p}}(\theta_i, \zeta_i)$$

And so, using equation (10) and rearranging terms:

$$\rho \int_{\theta_i,\zeta_i} [q(\bar{p} + \bar{m}\tau; \theta_i\,, \zeta_i) - q(\bar{p} + \bar{m}\tau; \theta_i\,, l)] dF_{\theta,\zeta}(\theta_i, \zeta_i)$$

$$= (1 - \rho) \int_{\theta_i} q(\bar{p}; \theta_i, h) dF_{\theta|p^s = \bar{p}}(\theta_i)$$

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

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

Using lemma 5 and plugging in yields:

$$
\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 5 and the CLD as in lemma 1:

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

In words, aggregate demand is contained between when all agents perceive a high price & have type h and when all agents perceive a low price & 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:

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

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:

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

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. $\square$

Claim: If $m_{it}\perp\beta_{it}$, $\beta_{it}\leq 0$ with probability one, and m_{it} has an exponential distribution, then:

$$\hat{m}\leq 0.5\Rightarrow\mathbb{E}[dwl_{it}|\tau]\leq-\frac{1}{2}\mathbb{E}[\tilde{\beta}_{it}]\tau^2\;\forall\tau$$

Proof: Since the variance of an exponentially-distributed random variable is its squared expected value:

$$\mathbb{E}(m^2_{it})=Var(m_{it})+(\mathbb{E}[m_{it}])^2=2(\mathbb{E}[m_{it}])^2$$

Also, from the independence of salience and sticker price responsiveness:

$$\hat{m} = \frac{\mathbb{E}[\tilde{\beta}_{it}]}{\mathbb{E}[\beta_{it}]} = \frac{\mathbb{E}[m_{it}]\mathbb{E}[\beta_{it}]}{\mathbb{E}[\beta_{it}]} = \mathbb{E}[m_{it}]$$

Combining these results and again using $m_{it} \perp \beta_{it}$ yields:

$$\hat{m} \leq 0.5 \Rightarrow \mathbb{E}[dwl_{it}|\tau] = -\frac{1}{2}\mathbb{E}[m_{it}^2]\mathbb{E}[\beta_{it}]\tau^2 = -\frac{1}{2}(2(\mathbb{E}[m_{it}])^2)\mathbb{E}[\beta_{it}]\tau^2 \leq -\frac{1}{2}\mathbb{E}[m_{it}]\tau^2$$

This inequality uses the fact that $\mathbb{E}[\beta_{it}] \leq 0 \leq \mathbb{E}[m_{it}]$. ■

A.3 Additional Proofs and Results from Section 4

Proof of Theorem 4: Pick any $\lambda_1 \neq 0$ and $\lambda_2 \in \mathbb{R}$. Pick a sequence $(\bar{p}_k, \tau_k)_{k=1}^{\infty}$ contained within the support of $(\bar{p}, \tau)$ that blows up in magnitude such that $\frac{\tau_k}{\bar{p}_k} \to \frac{\lambda_2}{\lambda_1}$. Then for each $k \in \mathbb{N}$, since price and consumption are observable, one can identify the distribution of:

$$\frac{q}{\bar{p}_k} = \frac{\alpha + \epsilon}{\bar{p}_k} + \beta + \tilde{\beta}\frac{\tau_k}{\bar{p}_k}$$

Taking the limit as $k \to \infty$ and multiplying both sides by λ_1, we can identify the distribution of:

$$\lambda_1 \lim_{k\to\infty} \frac{q}{\bar{p}_k} = \lambda_1 \left[\lim_{k\to\infty} \frac{\alpha + \epsilon}{\bar{p}_k} + \beta + \tilde{\beta}\frac{\tau_k}{\bar{p}_k}\right]$$

Since $\bar{p}_k$ blows up in magnitude as $k \to \infty$:

$$\lambda_1 \lim_{k\to\infty} \frac{q}{\bar{p}_k} = \lambda_1\beta + \lambda_2\tilde{\beta}$$

Note that we can alternatively pick a sequence $(\bar{p}_k, \tau_k)_{k=1}^{\infty}$ contained within the support of $(\bar{p}, \tau)$ so that $\bar{p}_k$ remains bounded while τ_k blows up. In this case, we can identify:

$$\lim_{k\to\infty} \frac{q}{\tau_k} = \lim_{k\to\infty} \frac{\alpha + \epsilon}{\tau_k} + \beta\frac{\bar{p}_k}{\tau_k} + \tilde{\beta} = \tilde{\beta}$$

Multiplying both sides by arbitrary $\lambda_2 \in \mathbb{R}$ yields:

$$\lambda_2 \lim_{k\to\infty} \frac{q}{\tau_k} = \lambda_2\tilde{\beta}$$

Thus, we can identify the distribution of $\lambda_1\beta + \lambda_2\tilde{\beta}$ for any $(\lambda_1, \lambda_2) \in \mathbb{R}^2$. By the Cramèr-Wold theorem, we can identify the joint distribution of $(\beta, \tilde{\beta})$, and so can identify $\mathbb{E}[dwl_i] = -\frac{1}{2}\mathbb{E}\left[\frac{\tilde{\beta}^2}{\beta}\right]\tau^2$ for any τ. ■

B. Empirical Application of Linear Model

We apply the linear specification of section 3.3 to data gathered by CLK (2009) on the aggregate consumption of beer in U.S. states between 1970 and 2003. First, we translate their model (in logs) to our linear specification. Second, we estimate the same equation of interest under different sets of controls:

$$y_{st} = \alpha + \beta\tau_{st}^{e} + \tilde{\beta}\tau_{st}^{s} + \gamma X_{st} + \epsilon_{st}$$

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 if $\overline{m} = 1$, i.e. agents never overreact to taxes. Results are presented in table 3. We also estimate various other specifications, again following CLK (2009), presented in table 4, meant to address concerns for spurious results – in particular, it could be the case that consumers react differently to the two tax rates as while sales taxes affect a variety of goods, excise taxes on beer affect only beer prices. The second last column of table 4 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. The share of ethanol people consume in the form of beer is insensitive to either tax rate.

	Baseline	Business cycle	Alcohol regulations	Region trends
Δ(excise tax)	-.08 (.03)	-.11 (.03)	-.1 (.03)	-.07 (.03)
Δ(sales tax)	-.05 (0.07)	-.02 (.07)	-.02 (.07)	-.03 (.07)
Δ(population)	-.0002 (0.0002)	-.0002 (.0002)	-.0001 (.0002)	-.0002 (.0002)
Δ(income per cap.)		.0002 (.00006)	.0002 (.00006)	.0002 (.00006)
Δ(unemployment)		-.09 (.03)	-.1 (.03)	-.09 (.03)
Alcohol reg. controls			X	X
Year FE	X	X	X	X
Region FE				X
$\hat{m}$	.67 (.87)	.2 (.62)	.23 (.65)	.46 (.99)
Sample size	1,607	1,487	1,487	1,487

Table 3: Estimating $\hat{m}$ with several sets of controls, following the specifications in Chetty et al. (2009) in the context of a linear model.

We repeat the exercise for Goldin and Homonoff (2013), who have a similar set-up with individual-level, cross-sectional data on cigarette consumption. The ratio of average responsiveness to sales taxes relative to the average responsiveness to excise taxes of about zero, although the estimate quite uncertain (see table 5 for details). Rosen (1976) uses a linear model, so we directly use his estimates from table 1 of his paper, reporting these results in table 6.

	Policy IV for excise tax	3-Year differences	Food exempt	Dep. var.: share of ethanol from beer
Δ(excise tax)	-.12 (.06)	-.24 (.1)	-.1 (.04)	.0003 (.0005)
Δ(sales tax)	-.02 (.07)	-.03 (.07)	-.05 (.07)	.001 (.001)
Δ(population)	-.0001 (0.0002)	-.002 (.0015)	-.00002 (.0002)	-.0000 (.0000)
Δ(income per cap.)	.0001 (.00006)	.0002 (.00007)	.0002 (.00007)	.0000 (.0000)
Δ(unemployment)	-.09 (.03)	-.03 (.03)	-.06 (.03)	-.0001 (.0004)
Alcohol reg. controls Year FE	X X	X X	X X	X X
$\hat{m}$	.17 (.54)	.11 (.3)	.52 (.68)	
Sample size	1,487	1,389	937	1,487

Table 4: Estimating $\hat{m}$ following the strategy of CLK (2009) in the context of a linear model. As in CLK, we use 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 substition with other alcoholic products, biasing the coefficient on excise tax relative to the one on sales tax. It shows that beer excise taxes have no discernable effect on the share of ethanol consumed from beer.

	Specification		
	1	2	3
Excise Tax	-4.85	-4.73	-4.87
	(.92)	(.92)	(.94)
Sales Tax	-.72	.42	.32
	(4.89)	(5.25)	(5.26)
Demographic controls	X	X	X
Econ. conditions controls		X	X
Income trend controls			X
State,year, and month FE	X	X	X
$\hat{m}$	.15	-.09	-.07
	(1.01)	(1.11)	(1.10)
Sample size	274,137	274,137	274,137

Table 5: 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).

	Hours/Year MTR at zero hours	Hours/Year MTR at full time	Hours/Week MTR at zero hours	Hours/Week MTR at full time
Wage	990.4	1218.1	18.8	21.96
	(74.88)	(106.3)	(1.75)	(2.48)
MTR*Wage	-950.7	-1480.7	-21.1	-26.58
	(269.7)	(324.5)	(6.29)	(7.58)
$\hat{m}$	0.96	1.21	1.12	1.21
	(.217)	(.175)	(.257)	(.227)
Sample size	2,545	2,545	2,545	2,545

Table 6: We report regression results directly from Table 1 of Rosen (1976).

CHAPTER 2

Tax Remittance Invariance

by Benjamin Glass and Joel Slemrod*

Abstract

In partial equilibrium, conventional wisdom holds that the impact of a sales tax does not depend on whether the buyer or the seller remits the tax to the government. We extend this result to a general setting. Equilibrium non-transfer activity depends only on the total tax liability, regardless of who bears the statutory remittance liability. In a competitive setting, this result applies when all agents must pay each other for sales and purchases. However, changing the remittance obligation may transfer wealth lump-sum, depending on legal production and consumption rights. Thus, the remittance neutrality of taxes is a variation on Coase's Theorem.

* University of Michigan

1. Introduction

All but lump-sum taxes affect production, trade, and consumption by altering incentives. Yet these incentives also depend on behavior-contingent transfers between agents. For instance, a seller of a good may have a legal obligation to remit a sales tax, but the buyer provides a transfer in the form of a price contingent upon purchase of the good. It is well understood that, if the market is competitive, then the price of the good would adjust in response to shifting the legal obligation to remit the tax to the buyer. This new equilibrium would yield the same amount of the good sold, as well as the same tax burden on each agent. In this way, it does not matter which agent remits the tax. Slemrod (2008) refers to this property as remittance invariance. We will use this term and remittance neutrality interchangeably.

We generalize this remittance neutrality result to a setting in which agents attempt to coordinate on taxed economic activity via action-contingent transfers. However, there is an exogenous network that determines the ability of agents to transfer to each other. This captures the intuition that some agents do not know each other, and so cannot pay each other for changing their own behavior. Yet, if the network of relationships between agents is connected, altering the remittance responsibility does not affect the non-transfer actions or tax incidence for any agents in the network. Intuitively, if the remittance responsibility shifts from agent A to agent B, agent B may require compensation from another agent, who demands compensation from yet another agent, until eventually agent A is required to compensate for the increased statutory burden on agent B.

In our setting, we assume a fixed payoff vector for failure to coordinate. This means the change in remittance obligations does not affect the bargaining power of any agents, ruling out income effects. In this environment, a redistribution of remittance obligations does not affect non-transfer actions or equilibrium payoffs. For example, one may interpret failure to coordinate as agents consuming their initial endowments. This requires that the tax code always lets one consume all initial endowments. This would not hold, for instance, if the sale of a good by agent A to agent B created a remittance obligation for unrelated third-party C. Then agent C might pay agent B (directly or through paying another agent to pay agent B) to not purchase the good from agent A. This yields the same substitution effect of the tax as if the remittance obligation were directly on agent B, but now the tax impoverishes agent C to the benefit of agent B.

Our result has strong similarities with Coase's Theorem (see Coase 1960), in which any configuration of well-defined ownership rights yields an efficient equilibrium in an environment with externalities. In either case, agents offer action-contingent transfers to modify each other's behavior. Coase emphasizes that efficiency requires agents to make contracts providing for action-contingent transfers. To the degree agents cannot contract with each other, externalities still yield inefficiency. We emphasize that a connected network in which any pair of connected agents can contract with each other has no contractibility issue.

Our model has agents strategically cooperating, but our main result also applies to competitive markets. Intuitively, the price mechanism is a coordination device, dictating to agents how to divide the surplus from their interactions. More formally, any Walrasian equilibrium is an allocation of resources such that no coalition of agents could all be better off with an alternative arrangement using their own endowed resources.[1] This means no group of agents could agree amongst themselves upon an alternative to the equilibrium allocation.

In Coase (1960), if agent A's activity hurts agent B, then the same result obtains whether agent A requires payment from agent B to engage in that activity *or* agent B must pay agent A to desist from the activity. But agents A and B are not indifferent as to who pays whom. Transferring property rights from A to B effectively transfers wealth from A to B. But this is the only effect of the change in property rights; it does not distort the behavior of either agent.

Similarly, if a producer and consumer's transactions generate a statutory tax burden on a third party, then there is a substantial transfer of wealth if the third party must pay off either the buyer or seller to *not* transact with each other. Formally, in a variant of our model in which agents pay each other to *not* buy or sell, one can redistribute wealth lump-sum corresponding with a change in the distribution of statutory tax liability to obtain the same non-transfer action profile and tax incidence when the total statutory tax liability for the same actions remains unchanged. We refer to this as an *impure* variation on remittance neutrality that includes income effects resulting from the change in the tax code.

Whether an alteration in remittance responsibility affects a lump-sum transfer depends in practice on legal details. For instance, suppose originally a customer remits taxes generated by purchases of a good using a credit card. Subsequently, the government shifts this remittance obligation onto the credit card company.[2] If the credit card company can demand a fee from

[1] In fact, there is no coalition of agents who could all be at least as well off, with some strictly better off. This is an extension of the First Welfare Theorem. One can find a demonstration of this point in Mas-Colell, Whinston, and Green (1995).
[2] In fact, Costa Rica requires credit card companies to remit taxes. See Brockmeyer and Hernandez (2016).

the customer for use of the credit card, then they will demand a fee equal to the tax obligation, and so the amount that the customer pays using the credit card remains unaltered. However, if the credit card company legally must permit the customer to use the credit card for these purchases, then instead the credit card company offers a rebate to the consumer equal to the size of the tax obligation abated by the consumer from reducing consumption from what it would have been had there been no tax.[3] This makes the consumer substantially wealthier than under the original remittance regime, yet the consumer faces the same opportunity cost of an additional unit of consumption. Lump-sum transferring funds from the credit card company to the consumer could then yield the original equilibrium level of consumption.

We are not aware of any tax code in which transactions yield tax obligations on third parties that were not involved in some way with the transaction. Yet we discuss the impure remittance invariance result for two reasons. One, as illustrated in the credit card example, a party nominally involved in a transaction may have no effective say in the transaction in practice. Two, even if we felt that no government currently imposed remittance obligations on third parties unrelated to the corresponding transaction, this would not imply that a government could not contemplate such a policy.[4]

In terms of the theoretical focus on tax remittances, this paper is most similar with Slemrod (2008), which discusses the conditions under which remittance responsibility affects outcomes

---

[3] This supposes two details. One, the credit card company has sufficient profit from other sources that it does not seek to (or otherwise cannot) avoid these payments by shutting down. Two, the company knows how much the consumer would consume in the absence of any taxes. If the company was not sure, it could choose a liberal estimate of this amount. That would yield a greater transfer of wealth from the company to the consumers.

[4] For instance, the Vickrey-Clark-Groves (VCG) mechanism generally requires payments from agents based on the reports (actions) of other players. See Mas-Colell, Whinston, and Green (1995), where they refer to the VCG mechanism as Groves-Clark. The VCG mechanism can be nested in our model with the mechanism designer as an agent connected to all other agents.

of economic significance. Slemrod argues that tax remittance responsibility matters to the degree it affects the administrative and compliance costs of taxation, as well as and tax evasion. Similarly, Kopczuk et al. (2016) provide evidence that diesel tax evasion is facilitated when (downstream) retailors have the remittance obligation compared to when (upstream) wholesalers do. But our model appears novel in considering the remittance problem beyond the bilateral interaction between supply and demand in a single market.

This paper also shares some theoretical similarities with Bagwell and Bernheim (1988). One, they study a connected network of dynasties, where connections in their setting indicate altruistic concern. Two, they use their connected network to demonstrate the robustness of economic decisions to initial conditions. However, Bagwell and Bernheim (1988) use a directional connected tree (parents giving to children), whereas we use an undirected connected network (i.e. any connected pair of people can transfer arbitrarily to each other). Also, Bagwell and Bernheim (1988) demonstrate an implausible neutrality to almost all public redistributions, taxes, and prices. In contrast, the result in this model still allows for the magnitude of aggregate tax burdens generated by activity to matter. It only precludes the distribution of remittance obligations from having anything more than a lump-sum effect.

We begin by demonstrating tax remittance invariance without specifying a specific market structure. Thereafter, we focus on competitive markets. We believe our main result is also applicable to imperfectly competitive markets, and so relates to the work of Weyl and Fabinger (2013), who demonstrate tax remittance invariance in various imperfectly competitive partial equilibrium market structures.

Section 2 develops the model, formalizing the economy as a one-shot game with an exogenous outside option for all players. Section 3 describes the main result, while section 4 maps this result to a competitive setting. Section 5 extends the competitive setting to account for different entitlements to transact. In section 6, we reinterpret the general model to consider tax administrative and compliance costs. Section 7 concludes.

2. Model

We describe the game in four stages. First, we lay out the tax environment. Second, we describe the exogenous connections between players. It is while describing these connections that we define what it means for the network of agents to be connected. Third, we describe the choice sets of agents. Finally, we define the payoff functions of the agents. It is only when we describe the payoff functions that the formal relationship between connections and transfers becomes clear.

First, let $\mathcal{N}$ denote the finite network of agents indexed by i, and let $t = 1, \dots, T$ denote the period. In each period t, agents i have non-transfer choice sets A_{it}. We think of these actions as production, trade, and consumption. An agent i may not have a meaningful choice in a period t, in which case the cardinality of A_{it} is one. Let $A_i \equiv \prod_{t=1}^{T} A_{it}$ denote the set of streams of non-transfer choices of agent i, and let $A \equiv \prod_{i \in \mathcal{N}} A_i$ denote the set of non-transfer action profiles. These action profiles yield tax remittance responsibility $\tau_i: A \to \mathbb{R}$ for each agent i. These taxes do not depend on transfers, ruling out taxes on transfers.[5]

[5] This precludes inheritance taxes and taxes on gifts. For instance, even if agent A had to remit one dollar to the government for every dollar given to agent B, agent A may still wish to transfer funds to agent B. However, that calculus changes dramatically if for every dollar agent A gives agent B, agent B must give that dollar to the government.

Second, agents are restricted in their transfer choices by their network connections. As we will make clear, these connections and the timing of the choice of transfer functions are exogenous. Because agents must mutually agree to an action-contingent transfer scheme, getting to choose the transfer function first has a strategic advantage. Formally, let $c_{ijt} \in \{0,1\}$ denote whether agent i is *connected* to agent j and able to choose the transfer function at time t. We assume that each agent i can make the choice of transfer function for agent j at most one time, and each agent can agree to the proposed transfer scheme. In other words, an offer is an ultimatum, not subject to renegotiation.

Assumption 1: For any $i, j \in \mathcal{N}$, if $\exists t: c_{ijt} = 1$, then:

(1) $c_{ijt'} = 0 \; \forall t' \neq t$

(2) $\exists t': c_{jit'} = 1$

Negotiations are one-shot, and every agent to whom terms are dictated has one opportunity to either take it or leave it. If $c_{ijt} = c_{jit} = 1$, then both agents must simultaneously agree at period t to an action-contingent transfer scheme. If $c_{ijt} = 0 \; \forall t$, then agent i cannot transfer to agent j or vice versa.

We assume that the network of agents is connected, meaning that there is a finite path of connections between all agents.

Assumption 2: For any agents i and i' there exists a sequence of agents $\{k_j\}_{j=1}^{L} \subseteq \mathcal{N}$, with $k_0 = i$ & $k_L = i'$, and a function $t(j) \in \{1, \dots, T\}$ such that:

$$\prod_{j=1}^{L} c_{k_{j-1},k_j,t(j)} = 1$$

In words, agent i can transfer to agent k_1, who can transfer to agent k_2, and so on to agent i'.

Third, we describe the strategy space. Letting $N \equiv |\mathcal{N}|$ denote the number of agents and $\mathcal{F}_t$ denote the set of functions mapping from A to $\mathbb{R}^N$ $\forall t$, each agent i in any period t chooses from $C_{it} \equiv A_{it} \times \mathcal{F}_t$. We let agents choose functions $g_{it}(a) \equiv \left(g_{ijt}(a)\right)_{j\in\mathcal{N}} \in \mathcal{F}_t$ for all agents in all periods for modeling simplicity. However, the payoff functions will clarify that the only choices of transfer functions g_{ijt} that matter are those for which $c_{ijt} = 1$. The interpretation is that agent $i \in \mathcal{N}$ offers in period $t \in \{1, \dots, T\}$ to pay $g_{ijt}(a)$ to agent $j \in \mathcal{N}$ conditional on action profile $a \in A$. This offer is also contingent on all connected agents agreeing on transfers. In every period t, all agents i choose from C_{it} knowing the choices of all agents in all previous periods.

Finally, we describe the payments of agents under two cases. To start, suppose that:

$$\forall i, j \in \mathcal{N}\ \forall t, t' \in \{1, \dots, T\}\ c_{ijt} = c_{jit'} = 1 \Rightarrow g_{ijt}(a) + g_{jit'}(a) = 0\ \forall a \in A$$

In words, this says that the agents agree on the transfer of wealth between each other. For instance, if agent i offers to pay ten dollars to agent j contingent on action profile a, then agent j agrees to receive ten dollars from agent i. If all agents agree, then each agent i receives payoff from action profile $a \in A$:

$$v_i\left(a, \sum_{j\in\mathcal{N}} c_{jit} g_{jit}(a) - \tau_i(a)\right)$$

Here v_i maps from the action profile and net increase in wealth for agent $i \in \mathcal{N}$ to the payoff for that agent i.[6] The payoff of an agent depends on the non-transfer action profile and the net tax-inclusive transfer of wealth.

Now suppose otherwise, i.e.:

$$\exists a \in A, i, j \in \mathcal{N}, t, t' \in \{1, \dots, T\}: c_{ijt} = c_{jit'} = 1, g_{ijt}(a) + g_{jit'}(a) \neq 0$$

This is to say that some pair of agents i and j fail to agree on transfers. In this case, each agent receives an outside option payoff of $\overline{u_i}$. In words, agents in this network engage in certain economic activity and transfers if and only if all agents in the economy agree to those activities and transfers, so that net transfers from agent i to agent j are opposite and equal to the net transfers from agent j to agent i. In this sense, every agent must cooperate for the market to function. For instance, in perfect competition all agents must be at least as well off as they would be without engaging the market.[7]

In short, letting $\mathbb{I}$ denote the indicator function mapping true statements to one and false statements to zero, payoff functions $V_i: ((C_{i't})_{i' \in \mathcal{N}})_{t=1}^{T} \to \mathbb{R}$ can be defined by:

$$\begin{aligned} &V_i(((a_{i't}, g_{i't})_{i' \in \mathcal{N}})_{t=1}^{T}) \\ &\qquad \equiv \mathbb{I}\big(c_{i'jt} = c_{ji't'} = 1 \nRightarrow g_{i'jt}(a) + g_{ji't}(a) \neq 0\big)\bar{u}_i \\ &\qquad + \mathbb{I}\big(c_{i'jt} = c_{ji't'} = 1 \Rightarrow g_{i'jt}(a) + g_{ji't}(a) = 0\big) v_i\left(a, \sum_{j \in \mathcal{N}} c_{jit} g_{jit}(a) - \tau_i(a)\right) \end{aligned}$$

[6] When we consider general equilibrium environments without transfers, we can assume that for any consumers i, v_i always maps $A \times [0, \infty)$ to greater values than to which it maps $A \times (-\infty, 0)$. Furthermore, for any $a \in A$ and $r > 0$, we will assume that $v_i(a, r) = v_i(a, 0)$ for such consumers i. Thus, agents never choose allocations that they cannot afford, and do not directly benefit from having leftover funds after all transactions occur. This all says that agents do not obtain direct utility from money, but must afford what they consume.

[7] This phrasing rules out lump-sum taxes. Alternatively, we could define $\overline{u_i}$ as the utility of agent i when not buying or selling, and yet paying lump-sum taxes. Then we would define τ_i as the net-of-lump-sum tax function.

Implicitly, we assume that the failure of some agents to negotiate precludes the other agents from attempting to coordinate amongst themselves. One can interpret the contract offers as all being contingent on the cooperation of all other agents, so that there is no meaningful communication between agents once some agents fail to agree. However, our results survive alterations to the non-cooperative payoff structure so long as one restricts attention to Nash equilibrium.[8]

In summary, this is a game with perfect information in which every agent $i \in \mathcal{N}$ in every period $t \in \{1, \dots, T\}$ chooses from C_{it}. All players i have payoff functions V_i, which depend on players' choices and exogenous connections. In the next section, we define equilibrium and provide a formal description of our headline result.

3. Main Result

Our principle finding pertains to the equilibrium notion of a subgame-perfect Nash equilibrium (SPNE). In words, the result holds that, if we replace one tax code with another tax code that yields the same aggregate tax burden for each action, then the same SPNE obtains. To state this result more formally, we first need to define strategies. Toward that end, let $C_t \equiv \prod_{i \in \mathcal{N}} C_{it}$ denote the set of all period t choice profiles, and let $H_t \equiv \prod_{t'=1}^{t-1} C_{t'}$ denote the set of histories possibly observed in period t, where $H_1 = \{\emptyset\}$. Let $H \equiv \bigcup_{t=1}^{T} H_t$ denote the set of all possible histories.

[8] Gale, Binmore, and Samuelson (1995) argue that agents playing the Ultimatum Game may learn to play Nash equilibria that are not subgame-perfect. In any case, our concern is to avoid a history with only some agents cooperating, where the actions of this set of cooperating agents yields tax burdens on non-cooperative agents. We also want to avoid the set of cooperating agents not being connected. We are unware of any real-world example of our theoretical concerns. We also believe our result would hold if $T = \infty$, and if agents can provide counteroffers.

For any $i, j \in \mathcal{N}$ and $t \in \{1, \dots, T\}$, a transfer strategy function $s_{ijt}^g: H \rightarrow \mathcal{F}_t$ satisfies $\forall \tilde{t} \geq t$ and $(a_{i't'}, g_{i't'})_{t'=1}^{\tilde{t}} \in H_{\tilde{t}}$:

$$s_{ijt}^g\left((a_{i't'}, g_{i't'})_{t'=1}^{\tilde{t}}\right) = g_{ijt}$$

In words, the transfer strategy function cannot assert past offers distinct from what the agent has already chosen. Agent i has already offered g_{ijt} to agent j in period $t \leq \tilde{t}$, and so cannot choose to have made any different offer to agent j in period t. Let $s_{it}^g(h) \equiv \left(s_{ijt}^g(h)\right)_{j \in \mathcal{N}}$ denote the period t transfer strategy of agent i $\forall h, i, t$. An action strategy $s_{it}^a: \mathcal{H} \rightarrow A_{it}$ also does not contradict the past, i.e. $\forall i, j, t, \tilde{t} \geq t$ and $(a_{i't'}, g_{i't'})_{t'=1}^{\tilde{t}} \in H_{\tilde{t}}$:

$$s_{it}^a\left((a_{i't'}, g_{i't'})_{t'=1}^{\tilde{t}}\right) = a_{it}$$

Let $s_{it}(h) \equiv \left(s_{it}^a(h), s_{it}^g(h)\right)$ denote a period t (total) strategy for agent i. Let $s_i(h) \equiv \left(s_{it}(h)\right)_{t=1}^T$ denote an overall strategy of agent i. Let $s(h) \equiv \left(s_i(h)\right)_{i \in \mathcal{N}}$ denote a strategy profile. Finally, it will prove convenient to let $s^a(h) \equiv \left(\left(s_{it}^a(h)\right)_{t=1}^T\right)_{i \in \mathcal{N}}$ denote the non-transfer action strategy profile and $s_{-i} \equiv (s_i')_{i' \neq i}$ denote the profile of strategies of agents who are not agent $i \in \mathcal{N}$.

Definition: A strategy profile s^* is a SPNE if for all agents i, for any history $h \in H$, and any strategy $s' \equiv \left(s_{i'}'\right)_{i' \in \mathcal{N}}$ with $s_{-i}' = s_{-i}^*$:

$$V_i(s^*(h); (\tau_{i'})_{i' \in \mathcal{N}}) \geq V_i(s'(h); (\tau_{i'})_{i' \in \mathcal{N}})$$

In words, an SPNE has all agents optimizing in response to any history.

Our main result states that, for any redistribution of tax remittance obligations such that the total tax obligation from an action profile a remains unchanged, there is an SPNE with the same equilibrium action profile and tax incidence. In the appendix, we constructively define a mapping from old to new equilibrium strategy profiles. The new SPNE provides the same financial incentives to all agents.

Theorem 1: *Suppose s^* is a SPNE of a game with tax functions τ_i $\forall i$. Let $g^*_{ij}(a)$ denote the corresponding equilibrium transfer functions. Now consider a new game with altered tax functions τ'_i such that:*

$$\sum_i \tau_i(a) = \sum_i \tau'_i(a) \ \forall a \in A$$

Then there exists a SPNE of this new game s'. Letting $g'_{ijt}(a)$ denote a new equilibrium transfer function, and $s^{a\prime}$ denote the new equilibrium non-transfer action strategy profile, the following hold $\forall a, i, h$:

$$\sum_{j \in \mathcal{N}} c_{jit} g'_{jit}(a) - \tau'_i(a) = \sum_{j \in \mathcal{N}} c_{jit} g_{jit}(a) - \tau_i(a)$$

$$s^{a\prime}(\emptyset) = s^a(\emptyset)$$

$$V_i\left(s'(h); \left(\tau'_{i'}\right)_{i' \in \mathcal{N}}\right) = V_i(s(h); (\tau_{i'})_{i' \in \mathcal{N}})$$

The idea of this proof is to consider an arbitrary connected network of agents. For such a network and for any action profile, one can find some agent $\kappa(1)$ who can be compensated by some agent $\lambda(1) \neq \kappa(1)$ for the change in the tax remittance obligation, such that the network of agents $\mathcal{N} \setminus \{\kappa(1)\}$ remains connected. One can continue this process so that, for any $n < N$, agents $\mathcal{N} \setminus \{\kappa(1), \ldots, \kappa(n)\}$ remains connected while $\lambda(n) \notin \{\kappa(1), \ldots, \kappa(n)\}$. We depict a connected network with such functions in figure 2.1. Because the aggregate tax burden of the action profile remains unchanged, the last agent $\kappa(N)$ also faces the same burden as before the tax remittance policy change. Thus, for any action profile, the net transfer to each agent remains unchanged. This leaves all agents with the same incentives as before the tax change.

4. Competitive Equilibrium

We now turn our focus towards the relationship between our main result and the competitive market environment most frequently studied in public economics. The standard remittance invariance result concerns changing the tax code between buyers and sellers of a good remitting the tax. Instead, we wish to apply our general result to a price-taking environment in which the tax code changes from one's transactions creating a remittance obligation for oneself to a new tax code in which the same transactions create that remittance obligation for someone else. This is as if before buyer A remitted the sales tax after purchasing from a seller B, but now that specific seller B had to remit the tax *even if* the buyer purchased the good from some other vender C.

An apparent tension arises from the cooperative setting of our main result, where any one agent can choose to end cooperation. In contrast, competitive equilibrium has a price mechanism that automatically coordinates the actions of agents. We resolve this difficulty by noting that one can consider prices to be set by an agent often referred to as a Walrasian auctioneer.[9] The Walrasian auctioneer, denoted by w, sets prices to minimize the sum of squared differences between supply and demand. Prices are then transfers between agents $i \neq w$ and w.

[9] As in Arrow and Debreu (1954).

Figure 2.1

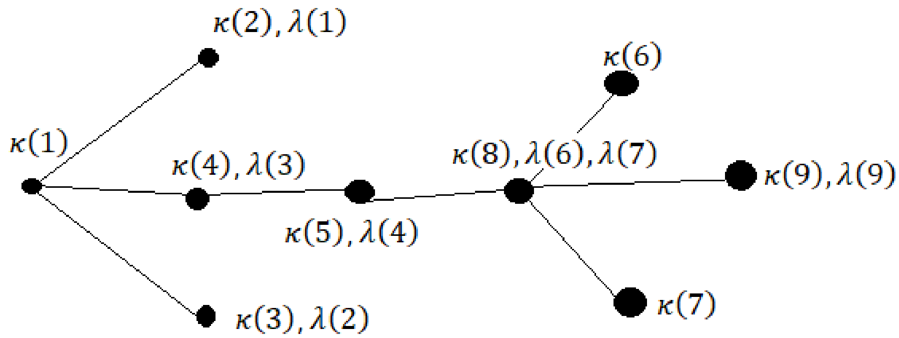


Figure 2.1: An example of a connected network with functions κ and λ, mappings from $\{1, \dots, N\}$ to $\mathcal{N}$, such that for any action profile, agents $\kappa(n)$ are compensated by agents $\lambda(n)$ so that they have the same net transfers as in the old equilibrium. We show in the appendix that κ is a bijection (with κ^{-1} a well-defined function on $\mathcal{N}$), while $\lambda(n)$ is connected to $\kappa(n)$ for $n = 1, \dots, N-1$. For completeness, we define $\lambda(N) \equiv \kappa(N)$. Note that λ is not a bijection, so that it's possible that $\lambda(n) = \lambda(n')$ for $n \neq n'$.

Another challenge comes from the limitations of the transfer function that agent w can use. These transfer functions must be linear in quantities bought and sold.[10] That the transfer functions are linear requires us to limit our attention to linear changes in the tax code. But more substantially, we must consider a greater domain on which these functions may be defined than is traditional in competitive equilibrium. This alteration is a resolution of a contractibility problem of statutory remittance obligation for the actions of others.

For example, consider an economy with one producer, one consumer, and some third agent. If there is a sales tax triggered by the sale of the good, then a standard competitive model has economic tax incidence borne by both the producer and the consumer. However, if the statutory burden of the sale of the good falls upon the additional agent, then the standard

[10] By linear, we mean transfers are additive and homogeneous of degree one with respect to quantities traded. This is as in the Arrow-Debreu model.

competitive model has the additional agent take the tax burden as given and immutable, whereas neither consumer nor producer internalize the burden created by their behavior.[11] Thus, shifting the statutory responsibility to the third party both alleviates the distortionary impact of the tax and shifts the economic tax incidence.

But suppose the third agent could charge the first two agents for their transaction of the taxed good. Then the third agent could charge an amount equal to the per unit tax rate, thus being made whole from the statutory obligation to remit the sales tax. Meanwhile, supposing the third agent charged the producer, the producer would deduct the price charged by the third agent from the price charged to the consumer to get the marginal revenue of selling a unit of the good. If the third agent instead charged the consumer, then the consumer would sum the charged prices to obtain the marginal cost of purchasing a unit of the good. Either way, the price charged by the third agent equals the sales tax, and so both economic activity and tax incidence remain unchanged.

A difficulty remains with the above story, depending on the setting. This seems quite plausible if the third agent were a credit card company whose role was essential to the transaction. But if the third party had no relationship with the transactions, then we may not think it plausible that the third agent can charge the first two agents for their commerce; perhaps instead the third agent should be paying them to *not* trade with each other. If so, then the change in the statutory liability will have created a lump-sum transfer from the third agent to the first two. All the same, if no income effects obtain for the producer or consumer, then

[11] Here we intend the "standard competitive model" to refer to a setting where agents face prices only for their own consumption, taking the actions of all other agents as given.

economic activity and tax incidences remain unchanged. More broadly, if the producer and consumer were forced to compensate the third agent via lump-sum payments that all agents take as given, then the imposition of the tax would have no impact on tax incidence or economic activity.

We address this point in the next section. For now, we lay out a competitive model in which all agents take prices for the economic decisions of all agents as given. We then demonstrate that a linear change in the statutory burden of taxes results in a new equilibrium in which both tax incidence and real economic activity remain the same.

Let L denote a finite number of goods. In the spirit of general equilibrium, let $z_{il}: A \to \mathbb{R}$ denote the excess demand function for good $l = 1, \dots, L$ from the choice of agent $i \in \mathcal{N}$.[12] Negative values of z_{il} denote net supply. We always allow agents to choose to not produce:

$$\forall i \; \forall a_{-i} \; \exists a_i \in A_i : z_i(a_i, a_{-i}) = \mathbf{0}$$

Here $\mathbf{0} \in \mathbb{R}^L$ denoting the zero vector. Market-clearing requires that:

$$\sum_{i \in \mathcal{N}} z_i(a) = \mathbf{0}$$

As a modeling convenience, we let $T = 2$. This merely captures the idea that first the Walrasian auctioneer sets prices, and then agents make their decisions. One agent is the Walrasian auctioneer, denoted w, who sets the prices such that the market clears. This agent seeks to minimize the sum of squared differences between supply and demand:

$$u_w = -\left\| \sum_{i \in \mathcal{N}} z_i(a) \right\|$$

[12] Note that z_i taking in arguments from A_{-i} permits externalities. Also, while we express our results in general equilibrium terms, our example at the end of the section makes clear that our analysis also applies for a single market in partial equilibrium.

But the Walrasian auctioneer cannot produce or consume any goods:

$$z_w(a) = \mathbf{0} \; \forall a \in A$$

We set $c_{wi1} = c_{iw2} = 1 \; \forall i \in \mathcal{N} \setminus \{w\}$, so the Walrasian auctioneer gets to set transfers first. Furthermore, $c_{ij1} = c_{ij2} = 0 \; \forall i, j \in \mathcal{N} \setminus \{w\}$. Thus, the transfers are exactly as determined by the Walrasian auctioneer. For simplicity, we suppress t subscripts, denoting g_{wi1} as g_{wi} and g_{iw2} as $g_{iw} \; \forall i \neq w$. Finally, there is a tax profile $\left(\tau_i(a)\right)_{i \in \mathcal{N}}$, where $\tau_w(a) = 0 \; \forall a$.

Suppose there is a SPNE of a game with the above structure yielding a non-transfer action profile a^* with vectors $p_{ij} \in \mathbb{R}^L \; \forall i, j \neq w$ such that $\forall a \in A$:

$$g_{iw}(a) = \sum_{j \neq i,w} p_{ij} * z_j(a) - \left[\sum_{j \neq i,w} p_{ji}\right] * z_i(a) + p_{ii} * z_i(a)$$

$$g_{iw}(a) + g_{wi}(a) = 0$$

$$g_{ij}(a) = 0 \; \forall i, j \neq w$$

In words, each agent $i \neq w$ accept linear tax-exclusive price vectors $\left(p_{ij}\right)_{j \neq w}$ imposed by the Walrasian auctioneer. This means each agent i transfers $p_{ij} * z_j(a)$ to agent j for net demand $z_j(a)$. In a standard Walrasian equilibrium, $p_{ij} = \mathbf{0} \; \forall i \neq j$, and there would be some vector p^* such that $p_{ii} = p^* \; \forall i \neq w$. Thus, p^* would denote the tax-exclusive price vector that agent i faces for choice of net demand $z_i(a)$. Here, we allow for far more flexibility, so that agents i and j can contract to avoid the transactions of agent j creating a remittance obligation for agent i. However, our results would survive assuming $\exists p^* = p_{ii} \; \forall i \neq w$.[13] Regardless, as in a standard Walrasian equilibrium, we assume the SPNE clears all markets:

[13] One can check that if $\exists p^* = p_{ii} \; \forall i \neq w$ in the old SPNE, then still $p_{ii} = p^*$ for all consumers i in the new SPNE.

$$\sum_i z_i(a^*) = \mathbf{0}$$

Finally, regarding the SPNE, we assume the outside option utility values are defined as:[14]

$$\bar{u}_i = \sup_{a_i : z_i(a_i, a^*_{-i}) = 0} v_i\left(a_i, a^*_{-i}, \sum_{j \neq i, w} \left[p_{ji} * z_i(a_i, a^*_{-i}) - p_{ij} * z_j(a_i, a^*_{-i})\right] - \tau_i(a_i, a^*_{-i})\right)$$

This says that agents can choose to consume only what they have, subject to taxes and their own prices for other transactions. While agents may well choose to benefit from negative values of $p_{ij} * z_j(a_i, a^*_{-i})$, it may appear excessively strong to suppose that a decentralized marketplace can require agents to pay positive values of $p_{ij} * z_j(a_i, a^*_{-i})$ contrary to the wishes of the agent i; compelling involuntary payment is generally considered the domain of the tax authority. We return to this point in the next section.

Consider a new remittance responsibility regime $\left(\tau'_i(a)\right)_{i \in \mathcal{N}}$ with $\tau'_w(a) = 0\ \forall a$ such that $\forall a \in A$:

$$\forall i \neq w\ \exists \Delta\tau_{ij} = \left(\Delta\tau_{ijl}\right)_{l=1}^{L}\ \forall j \neq w : \tau'_i(a) - \tau_i(a) = \sum_{j \neq w} \left[\Delta\tau_{ij} * z_j(a) - \Delta\tau_{ji} * z_i(a)\right]$$

This says that remittance liabilities have been rearranged amongst buyers and sellers of goods, but the overall tax liability resulting from the same aggregate production and consumption remains unchanged:

$$\sum_i \tau_i(a) = \sum_i \tau'_i(a)\ \ \forall a \in A$$

The rearranging of statutory liability is linear in net demand to maintain linear pricing.

---

[14] One may note that, contrary to standard general equilibrium analysis, we allow for net payments to directly enter the utility function. One might have some value $\underline{u}$ such that $\forall i\ u_i(a, m) \geq \underline{u} \Leftrightarrow m \geq 0$, so that no agent can ever choose to owe more money than they are worth.

We say that the new tax schedule $(\tau_i')_{i\in\mathcal{N}}$ with the same values $\bar{u}_i$ $\forall i$ yields an equilibrium with identical economic activity and tax incidence if it yields a new SPNE with the same non-transfer action profile a^* and new (tax-exclusive) prices $p_{ij}' \in \mathbb{R}^L$ $\forall i, j \neq w$ such that:

$$g_{iw}'(a) = \sum_{j\neq i,w} p_{ij}' * z_j(a) - \left[\sum_{j\neq i,w} p_{ji}'\right] * z_i(a) + p_{ii} * z_i(a)$$

$$g_{iw}'(a) + g_{wi}'(a) = 0$$

$$g_{ij}'(a) = 0 \; \forall i, j \neq w$$

In words, the prices adjust so that each agent has the same budget constraint as they would have faced if the tax code had not changed.

Theorem 2: There is a new equilibrium with transfers from the new tax code $\left(\tau_i'(a)\right)_{i\in\mathcal{N}}$ yielding the same real economic outcomes and tax incidence with:

$$p_{ij}' \equiv p_{ij} - \Delta\tau_{ij} \; \forall i, j \neq w: i \neq j$$

$$p_{ii}' \equiv p_{ii} \; \forall i \neq w$$

Proof: Note that one can write the statutory burden on the Walrasian auctioneer in the initial equilibrium as:

$$\tau_w(a) = 0 = \sum_{i\neq w} [\tau_i'(a) - \tau_i(a)] = \sum_{i\neq w}\sum_{j\neq w} \Delta\tau_{ij} * z_j(a)$$

Then, by the previous theorem, there is a new SPNE from the tax profile $\left(\tau_i'(a)\right)_{i\in\mathcal{N}}$ with the same non-transfer action profile a^*, so that $\sum_i z_i(a^*) = \mathbf{0}$, and such that for any $i \neq w$ and any $a \in A$:[15]

$$g_{iw}'(a) = \sum_{j\neq i,w} \left[p_{ij} + \Delta\tau_{ij}\right] * z_j(a) - \sum_{j\neq i,w} \left[p_{ji} + \Delta\tau_{ji}\right] * z_i(a) + p_{ii} * \mathrm{z_i}(a)$$

$$g_{iw}'(a) + g_{wi}'(a) = 0$$

$$g_{ij}'(a) = 0 \ \forall j \neq w$$

Furthermore, this SPNE yields the same outside option utility as in the old equilibrium $\forall i$:

$$\sup_{a_i : z_i(a_i, a_{-i}^*)=0} \mathcal{v}_i\left(a_i, a_{-i}^*, \sum_{j\neq w} \left[p_{ji}' * z_i(a_i, a_{-i}^*) - p_{ij}' * z_j(a_i, a_{-i}^*)\right] - \tau_i'(a_i, a_{-i}^*) \right)$$

$$= \sup_{a_i : z_i(a_i, a_{-i}^*)=0} \mathcal{v}_i\left(a_i, a_{-i}^*, \sum_{j\neq w} \left[\left[p_{ji} - \Delta\tau_{ji}\right] * z_i(a_i, a_{-i}^*) - \left[p_{ij} - \Delta\tau_{ij}\right] * z_j(a_i, a_{-i}^*)\right]\right.$$

$$\left. - \tau_i(a_i, a_{-i}^*) - \sum_{j\neq w} \left[\Delta\tau_{ij} * z_j(a) - \Delta\tau_{ji} z_i(a)\right] \right)$$

$$= \sup_{a_i : z_i(a_i, a_{-i}^*)=0} \mathcal{v}_i\left(a_i, a_{-i}^*, \sum_{j\neq w} \left[p_{ji} * z_i(a_i, a_{-i}^*) - p_{ij} * z_j(a_i, a_{-i}^*)\right] - \tau_i(a_i, a_{-i}^*) \right) = \bar{u}_i$$

Thus, defining $p_{ij}' \equiv p_{ij} + \Delta\tau_{ij}\ \forall i, j \neq w$ yields the desired result. ■

As a concrete example, consider agents $\mathcal{N} \equiv \{1,2,3,w\}$ with $L = 1$. We can think of agent 1 as a seller, agent 2 as a buyer, agent 3 a third party such as a credit card company. Formally, $A_1 = A_2 = \{0,1\}$, $A_3 = A_w = \{0\}$, and $c_{iw1} = c_{wi2} = 1\ \forall i = 1,2,3$, $c_{ijt} = 0$ otherwise.

[15] This claim also uses the fact that the choices of g_{ijt} are payoff-irrelevant when $c_{ijt} = 0$, so without loss of generality one can adjust those values so that g_{ijt} is always the same function.

Let $z_1(a) = -1$ if and only if $a_1 = 1$, $z_2(a) = 1$ if and only if $a_2 = 1$, and $z_3(a) = z_w(a) = 0\ \forall a \in A$. In words, agent 1 can decide to sell one unit to agent 2, and agent 2 can decide to buy one unit from agent 1, where prices are moderated by the Walrasian auctioneer. Suppose originally there is no tax, yielding a trading equilibrium action profile a^* so that:

$$a_1^* = a_2^* = 1$$

In words, agent 1 sells one unit of the good to agent 2. Suppose for simplicity that $p_{11} = p_{22}$ and otherwise $p_{ij} = 0$, so that the only exchange of funds is an amount $p_{11} = p_{22}$ transferred (via the Walrasian auctioneer w) from agent 1 to agent 2 for the sale of one unit of the good. Note that agent 3 does not demand any payments from the trading agents, reflecting the absence of any statutory remittance obligation generated by the trade. In fact, agent 3 makes no meaningful decisions at all.

Now, suppose that the transfer of the good yields a tax burden of $\Delta\tau_{11} = 10$ and $\Delta\tau_{31} = -10$. In words, trading the good causes the third party to pay the government ten dollars, while causing the government to pay the seller of the good ten dollars.[16] Then setting $p'_{31} = 10$ and $p'_{ij} = p_{ij}$ otherwise yields the same trading equilibrium a^* and the same values $V_i\ \forall i$. Now the third party demands payment from the seller of the ten dollars that the trade obligates the third party to remit. This realigns the incentives of all agents as in the original equilibrium.

[16] Recall that selling the good means $z_1 = -1$. This means that selling the good yields a remittance obligation for agent 1 of $10 * -1 = -10$, i.e. the government pays agent 1 ten dollars.

5. Variation in Entitlements

In the previous section, we permitted agents to demand payments from each other for each other's actions at prices set competitively. This may not be realistic. For instance, if a consumer purchase from a producer induces a statutory burden on a third party, one might imagine the third party paying the consumer to *not* purchase the good. In a sense, this variation transfers the entitlement to decide purchases entirely onto the consumer. We now explicitly model this process, finding that this redefinition of rights effectively transfers wealth towards the consumer. Different rules for determining choice sets that affect a change in statutory remittance obligation contain a lump-sum tax component, so that a transfer of wealth between agents corresponding with the change in statutory obligations is both necessary and sufficient to make the new tax code yield the same non-transfer action profile and utility levels for all agents. This is analogous to Coase's Theorem, where the distribution of property rights affects the distribution of utilities, but not efficiency. This is not remittance invariance as usually intended, but rather an invariance result with income effects.

For simplicity, we say that the original equilibrium has $p_{ij} = \mathbf{0}\ \forall i \neq j$. But now the new equilibrium has values $\bar{z}_{ij} = \left(\bar{z}_{ijl}\right)_{l=1}^{L}\ \forall i, j \neq w$ such that each agent taking prices p'_{ij} as given maximizes:

$$u_i\left(a, \sum_{j\neq i} p'_{ji} * \left[z_i(a) - \bar{z}_{ji}\right] - \sum_{j\neq i} p'_{ij} * \left[z_j(a) - \bar{z}_{ij}\right] - p_{ii} * z_i(a) - \tau'_i(a)\right)$$

The vectors $\bar{z}_{ij}$ denote baseline values of net demand to which agent j is entitled without payment to agent i. The amount that agent i pays agent j is then a linear function of the deviation of agent j's net demand from this baseline. This contains our model from the previous section, specifically when $\bar{z}_{ij} = \mathbf{0}\ \forall i, j$. But as $\bar{z}_{ijl}$ increases for l such that $p'_{ijl} > 0$, agent i receives less money from agent j. If $\bar{z}_{ij} > z_j(a)$ and $p'_{ij} > \mathbf{0}$, then agent i is then paying agent j for not increasing net demand.[17]

If $p'_{ij} = -\Delta\tau_{ij}\ \forall i, j \neq w$ as in the previous section, then each agent $i \neq w$ would maximize:

$$u_i\left(a, \sum_{j \neq i} \Delta\tau_{ij} * \bar{z}_{ij} - \sum_{j \neq i} \Delta\tau_{ji} * \bar{z}_{ji} - p_{ii} * z_i(a) - \tau_i(a)\right)$$

Note that now the tax and corresponding price changes the income of each agent $i \neq w$ by $\sum_{j \neq i} \Delta\tau_{ij} * \bar{z}_{ij} - \sum_{j \neq i} \Delta\tau_{ji} * \bar{z}_{ji}$. In words, the change in statutory remittance obligations can have a lump-sum transfer impact because the change in the tax code now changes the wealth of agents even in the absence of any market transactions. If there were no income effects, then there could be a new equilibrium from the tax code $(\tau'_i)_{i \in \mathcal{N}}$ with the same economic tax incidence and non-transfer action profile. More generally, define lump-sum transfers:

$$T_i \equiv \sum_{j \neq i} \Delta\tau_{ji} * \bar{z}_{ji} - \sum_{j \neq i} \Delta\tau_{ij} * \bar{z}_{ij}\ \ \forall i \neq w$$

For completeness, define $T_w \equiv 0$.

Corollary: A new tax schedule $(\tau'_i + T_i)_{i \in \mathcal{N}}$ has the same aggregate statutory obligation as $(\tau_i)_{i \in \mathcal{N}}$. Furthermore, there is a new equilibrium with the same real economic activity and tax incidence as the old equilibrium.[18]

Proof: Defining $p'_{ij} = \Delta\tau_{ij}\ \forall i, j \neq w$ yields the same optimization problem for all agents. Note that $\sum_i T_i = 0$. ■

[17]Put differently, if $\forall l\ \bar{z}_{ijl} \geq z_{jl}(a)$ and $p'_{ijl} > 0$, with $\bar{z}_{ij} \neq z_j(a)$, then $p'_{ij} * \left[z_j(a) - \bar{z}_{ij}\right] < 0$.

[18] Formally, $\left(\tau'_i + T_i\right)(a) = \tau'_i(a) + T_i\ \forall a \in A\ \forall i$.

For concreteness, consider the example from the end of section 4, but now let $\bar{z}_{31} = -1$. This means agent 3 must pay agent 1 to *not* make the trade, rather than agent 1 paying agent 3 for the right to make the trade. Then setting $T_1 = -10$, $T_2 = 0$, and $T_3 = 10$, with $p'_{31} = 10$ and $p'_{ij} = p_{ij}$ otherwise yields trade in the new equilibrium. In words, the government first transfers ten dollars from agent 1 to agent 3. Then, agent 3 pays agent 1 ten dollars if agent 1 does not sell the good. However, because this provides all agents the same incentives as the original equilibrium, agent 1 still sells the good to agent 2 in the new equilibrium.

6. Tax Administrative and Compliance Costs

Finally, we want to introduce considerations of imperfections in the tax collection technology. We do so to formalize a dual variant of the claim in Slemrod (2008) that the government should structure remittance obligations to minimize compliance and administrative costs. The most straightforward way to formalize and demonstrate this point is to reinterpret the model, so that now $\tau_i(a)$ denotes the cost to agent $i \in \mathcal{N}$ of the remittance responsibility generated by non-transfer action profile $a \in A$. We now refer to τ_i as the *remittance cost function* for agent $i \in \mathcal{N}$.

The government chooses the profile of tax functions $(\tau_i)_{i \in \mathcal{N}}$ from some family $\mathcal{T}$ of remittance cost functions, subject to a revenue correspondence $R: \mathcal{T} \times A \rightrightarrows \mathbb{R}$. The correspondence $R(\tau, a)$ is the amount of money that the government knows it can obtain from non-transfer action profile a and profile of remittance cost obligations $\tau(a)$, netting out tax administrative costs. There is some function $V^{gov}: \mathbb{R}^N \times \mathbb{R} \to \mathbb{R}$ that the government maximizes subject to $\mathcal{T}$ and R:

$$(\tau^*, r^*) \in \underset{\substack{\tau\in\mathcal{T}, r\in R(\tau, s^a(\emptyset)) \\ s \text{ is an SPNE for } \tau}}{\operatorname{argmax}} V^{gov}\left(\left(V_i(s(\emptyset);\tau)\right)_{i\in\mathcal{N}}, r\right)$$

In words, the government policy to maximize its objective, understanding that agents will choose an SPNE thereafter.[19]

This lets us make our formal claim. Suppose the government optimally chooses (τ^*, r^*) as previously described. Then, taking the resulting SPNE non-transfer action profile $s^a(\emptyset)$ as given, the government chooses to maximize revenue conditional on the non-transfer actions of agents and the aggregate remittance cost obligation.

Proposition: If V^{gov} is strictly increasing in its last argument, then for any solution (τ^*, r^*) to the government's optimization problem:

$$r^* \in \underset{\substack{\tau\in\mathcal{T}:\ \sum_i \tau_i(a)=\sum_i \tau_i^*(a)\ \forall a\in A \\ s \text{ is an SPNE for } \tau^*}}{\operatorname{argmax}} R\left(\tau, s^a(\emptyset)\right)$$

Proof: For some SPNE s^* with corresponding non-transfer action profile s^{a*}, the choice of (τ^*, r^*) must satisfy:

$$r^* \in \underset{\tau\in\mathcal{T}, r\in R(\tau, s^{a*}(\emptyset))}{\operatorname{argmax}} V^{gov}\left(\left(V_i(s^*(\emptyset);\tau)\right)_{i\in\mathcal{N}}, r\right)$$

Since V^{gov} is strictly increasing in its last argument:

$$r^* \in \underset{\substack{\tau\in\mathcal{T}:\ V_i(s(\emptyset);\tau)=V_i(s^*(\emptyset);\tau^*) \\ s \text{ is an SPNE for } \tau}}{\operatorname{argmax}} R\left(\tau, s^a(\emptyset)\right)$$

[19] Note that the formalism allows for the government to choose the SPNE. This would be in line with the mechanism design and principal-agent literatures. We impose the assumption to avoid issues of equilibrium selection in the government optimization problem.

The desired result follows from Theorem 1. ■

For a concrete example, suppose for every dollar that agent i is obligated by the tax code to pay, it costs that agent ϕ_i^c dollars total to remit that dollar to the government, i.e. it costs each agent i a compliance cost of $\phi_i^c - 1$ to then remit one dollar to the government. In addition, the government must pay a tax administrative cost of $\phi_i^a - 1$ to process ϕ_i^a dollars, so that agent i must remit ϕ_i^a dollars for the government to increase revenue, net of tax administrative costs, by one dollar. Thus, each agent i who bears a remittance cost obligation of $\phi_i^a \phi_i^c$ remits ϕ_i^a to the government, who pays tax administrative costs to be left with one dollar. This yields a definition of a revenue function:[20]

$$R(\tau, a) \equiv \sum_{i \in \mathcal{N}} \frac{\tau_i(a)}{\phi_i^a \phi_i^c}$$

From the previous proposition, we know that any solution (τ^*, r^*) to the government's optimization problem with corresponding SPNE action profile s^{a*} has:

$$\tau_i^*\left(s^{a*}(\emptyset)\right) = 0 \ \forall \phi_i^a \phi_i^c > \min\left\{\phi_{i'}^a \phi_{i'}^c\right\}$$

In words, the government requires remittances only from the one agent for whom the resulting remittance cost obligation is minimal.

7. Conclusion

We formalize the conditions under which, in an environment without administrative or compliance costs of taxation or tax evasion, only aggregate statutory tax obligations matter for real economic behavior; the distribution of statutory remittance obligations does not affect

[20] We assume that $\phi_i^a, \phi_i^c > 0 \ \forall i$. If we wanted, we could keep R a correspondence with $R(\tau, a) \equiv \left[0, \sum_{i \in \mathcal{N}} \frac{\tau_i(a)}{\phi_i^a \phi_i^c}\right]$, at least so long as the government does not remit funds to consumers, i.e. $\tau_i(a) \geq 0 \ \forall i \ \forall a$.

these outcomes. There can be lump-sum components to changing remittance obligations, but not when all agents must agree to any transactions.

One might consider objectives regarding the procedure of how taxes are collected, independent of the welfare impact these tax remittance obligations have on agents. Kaplow and Shavell (2001) argue that such considerations of horizontal equity necessarily make a social planner prefer some outcomes over Pareto-superior alternatives. This result largely precluded future work on horizontal equity, but we think our result suggests hope for future work in this area. After all, any restructuring of the tax code that increases some notion of “fairness” while leaving the aggregate remittance obligation of any action profile (as well as tax administrative costs and incentives to evade taxes) unchanged yields the same tax revenue.   This suggests one might separate fairness and efficiency considerations analogously to the traditional distinction between concerns for equity and efficiency.

Future work should also focus on the effect of altering remittance obligations on tax administrative and compliance costs. In addition, none of these results account for information asymmetries. Surely imposing a tax remittance obligation on retailers has a distinct effect from imposing this obligation on consumers if consumers will not notice taxes excluded from sticker prices, as in Chetty et al. (2009), yet would notice taxes that they must pay.

References

Arrow, Kenneth J., and Gerard Debreu. "Existence of an Equilibrium for a Competitive Economy." *Econometrica: Journal of the Econometric Society* (1954): 265-290.

Bernheim, B. Douglas, and Kyle Bagwell. "Is Everything Neutral?." *Journal of Political Economy* 96.2 (1988): 308-338.

Brockmeyer, Anne; Hernandez, Marco. 2016. "Taxation, Information, and Withholding: Evidence from Costa Rica." *Policy Research Working Paper; No. 7600. World Bank, Washington, DC. World Bank. https://openknowledge.worldbank.org/handle/10986/24140 License: CC BY 3.0 IGO."*

Coase, R. H. "The Problem of Social Cost." *JL & Econ.* 3 (1960): 1-44.

Chetty, Raj, Adam Looney, and Kory Kroft. "Salience and Taxation: Theory and Evidence." *American Economic Review*99.4 (2009): 1145-77.

Kaplow, Louis, and Steven Shavell. "Any Non-Welfarist Method of Policy Assessment Violates the Pareto Principle." *Journal of Political Economy* 109.2 (2001): 281-286.

Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod. 2016. "Does Tax-Collection Invariance Hold? Evasion and the Pass-Through of State Diesel Taxes." *American Economic Journal: Economic Policy*, 8 (2): 251-86.

Mas-Colell, Andreu, Michael Dennis Whinston, and Jerry R. Green. *Microeconomic Theory*. Vol. 1. New York: Oxford university press, 1995.

Slemrod, Joel. "Does it Matter who Writes the Check to the Government? The economics of tax Remittance." *National Tax Journal* (2008): 251-275.

Weyl, E. Glen, and Michal Fabinger. "Pass-through as an Economic Tool: Principles of Incidence under Imperfect Competition." *Journal of Political Economy* 121.3 (2013): 528-583.

Appendix for Chapter 2

Here we demonstrate the main result of this paper. To do so, we must first prove the lemma described in the intuition from section 3. This lemma is a result from graph theory, so we use notation from graph theory to express the result.

Definition: For any finite set V, for any set E of two-element subsets of V, the (undirected) graph $G = (V, E)$ is *connected* if for every distinct $v, \tilde{v} \in V\ \exists \{v_k\}_{k=1}^{K}$,for some finite $K > 1$ such that defining $e_k \equiv \{v_k, v_{k+1}\}$ for $k = 1, \dots, K-1$ yields:

$$e_k = e_{k'} \Rightarrow k = k'$$

$$v_0 = v, v_K = \tilde{v}$$

$$e_k \in E\ \forall k = 1, \dots, K-1$$

In words, a graph is connected if every two distinct vertices have a path connecting them. Our lemma is that for any connected graph of multiple vertices, we can always remove any one of at least two of the vertices and still have a connected graph.

Lemma: If $G = (V, E)$ is connected and $|V| > 1$, then there are at least two distinct points $v_1, v_2 \in V$ such that for $n = 1,2$, the induced subgraph from removing point v_n, denoted $G[V \setminus \{v_n\}]$, is also connected.[21]

[21] Given graph G, then $G[V \setminus V'] \equiv \left(V \setminus V', E \cap \left((V \setminus V') \times (V \setminus V')\right)\right)$ for any $V' \subset V$.

Proof: We prove this by induction on the cardinality of $|V|$. If $|V| = 2$, the claim is trivial. Suppose the claim holds up to cardinality $|V| - 1$. Pick arbitrary $v^* \in V$. One can show that there are disjoint subsets V_k for $k = 1, \dots, K$, with $K < |V|$, such that $v \in V_k$, $v' \in V_{k'}$, and $k \neq k'$ implies that no path connects v and v', yet:[22]

$$\bigcup_{k=1}^{K} V_k = V \setminus \{v^*\}$$

One can pick points v_1' and v_2' such that $G(V_1 \setminus \{v_n'\})$ is connected for $n = 1,2$. If $\{v_n', v^*\} \notin E$ for $n = 1,2$, then we can define $v_n \equiv v_n'$ for $n = 1,2$. If $\{v_n', v^*\} \in E$ for $n = 1,2$, then we can similarly define $v_n \equiv v_n'$ for $n = 1,2$. Finally, assume there is one point, say v_2', connected to v^* by a path, then define $v_1 \equiv v_1'$. If in this case $K = 1$, then define $v_2 \equiv v^*$. Otherwise, we can perform the same exercise with V_2 to get v_2. ■

Proof of Theorem: First, we want to map our notion of a connected network of agents into the definition of a connected graph. Formally, defining $E \equiv \left\{\{i,j\} \mid i \neq j, \exists t\colon c_{ijt} = 1\right\}$, we want to show that the graph $(\mathcal{N}, E)$ is connected. Toward that end, take any $i, j \in \mathcal{N}$. Then by assumption, there is a finite sequence $\{k_l\}_{l=0}^{L}$ and function $t(l) \in \{1, \dots, T\}$, such that $k_0 = i$, $k_L = j$, and:

$$c_{k_{l-1}k_l t(l)} = 1 \ \forall l = 1, \dots, L$$

We inductively define a function:

$$\tilde{l}(0) \equiv \max_{k_{l'} = i} l'$$

[22] One can also demonstrate this claim inductively on the cardinality of $|V|$. The claim is trivial for $|V| = 2$. Supposing it holds up to cardinality $|V| - 1$, adding one additional point $\tilde{v}$ means that the union of $\{\tilde{v}\}$ with all old subsets that have points connected to $\tilde{v}$ via a path form one such set V_1, whereas the other old subsets remain unaltered.

$$\tilde{l}(l) \equiv \max_{k_{l'}=k_{\tilde{l}(l-1)+1}} l' \;\; \forall l = 1, \ldots, L$$

Define $\tilde{L} \equiv \min_{\tilde{l}(l)=L} l$; this value exists and $\tilde{l}$ is invertible because $\tilde{l}(l)$ maps to $\{1, \ldots, L\}$ and is strictly increasing. Thus, we can define:

Note that defining $v_l = k_{\tilde{l}(l)}$, the sequence of points $\{v_l\}_{l=0}^{\tilde{L}} \subseteq \mathcal{N}$ such that:

$$v_0 = k_{\max_{k_{l'}=i} l'} = i$$

$$v_{\tilde{L}} = k_{\tilde{l}\left(\min_{\tilde{l}(l)=L} l\right)} = k_L = j$$

$$\{v_l, v_{l+1}\} = \left\{k_{\tilde{l}(l)}, k_{\tilde{l}(l+1)}\right\} = \left\{k_{\tilde{l}(l)}, k_{\max_{k_{l'}=k_{\tilde{l}(l)+1}} l'}\right\} = \left\{k_{\tilde{l}(l)}, k_{\tilde{l}(l)+1}\right\} \in E \;\forall l = 1, \ldots, \tilde{L} - 1$$

In addition, suppose for values $l, l' \in \{1, \ldots, \tilde{L} - 1\}$ we had:

$$\{v_l, v_{l+1}\} = \{v_{l'}, v_{l'+1}\}$$

That would imply:

$$v_{l+1} = v_{l'+1} \Rightarrow k_{\tilde{l}(l+1)} = k_{\tilde{l}(l'+1)}$$

From the definition of $\tilde{l}$:

$$v_l = v_{l'} \Rightarrow \tilde{l}(l+1) = \tilde{l}(l'+1)$$

Finally, since $\tilde{l}$ is strictly increasing:

$$l = l'$$

Thus, we have that $\forall l, l' \in \{1, \ldots, \tilde{L}\}$:

$$\{v_l, v_{l+1}\} = \{v_{l'}, v_{l'+1}\} \Rightarrow l = l'$$

We conclude that $\{v_l\}_{l=0}^{\tilde{L}}$ is a path connecting i to j. Since i and j were arbitrarily chosen, $(\mathcal{N}, E)$ is connected.

Second, we can inductively define κ and λ, confirming in the process that they satisfy properties sufficient for describing a compensating transfer scheme. To start, the lemma indicates that $\exists\kappa(1) \in \mathcal{N}$ such that $G[\mathcal{N} \setminus \{\kappa(1)\}]$ is connected. Since $(\mathcal{N}, E)$ is connected, $\exists\lambda(1): \{\kappa(1), \lambda(1)\} \in E$. This means that:

$$\exists t: c_{\lambda(1)\kappa(1)t} = 1$$

$$\kappa(1) \neq \lambda(1)$$

Now suppose by inductive hypothesis that for some value $n \in \{1, \dots, N-2\}$, we have already defined $\forall n' = 1, \dots, n$ values $\kappa(n')$ and $\lambda(n')$ such that, defining $\mathcal{N}_{n'} \equiv \{\kappa(1), \dots, \kappa(n')\}$:

$$\exists t: c_{\lambda(n')\kappa(n')t} = 1$$

$$\lambda(n') \notin \mathcal{N}_{n'}$$

That furthermore:

$$\forall n', n'' \in \{1, \dots, n\}, \kappa(n') = \kappa(n'') \Rightarrow n' = n''$$

And finally, the graph $G[\mathcal{N} \setminus \mathcal{N}_n]$ is connected. We can define $\kappa(n+1) \in \mathcal{N} \setminus \mathcal{N}_n$ such that the graph $G[\mathcal{N} \setminus \mathcal{N}_{n+1}]$ is connected. Since $G[\mathcal{N} \setminus \mathcal{N}_n]$ is connected, $\exists\lambda(n+1) \in \mathcal{N} \setminus \mathcal{N}_n$ such that:

$$\exists t: c_{\lambda(n+1)\kappa(n+1)t} = 1$$

$$\kappa(n+1) \neq \lambda(n+1)$$

Thus, $\lambda(n+1) \notin \mathcal{N}_n \cup \{\kappa(n+1)\} = \mathcal{N}_{n+1}$. Finally, since $\kappa(n+1) \notin \mathcal{N}_n$:

$$\forall n'' \in \{1, \dots, n, n+1\}, \kappa(n+1) = \kappa(n'') \Rightarrow n+1 = n''$$

We can conclude that all claims continue to hold with the definitions of κ and λ extended to the domain $\{1, \dots, n+1\}$. By induction, we can define κ and λ on $\{1, \dots, N-1\}$ satisfying the above claims. We complete the definition of κ and λ on $\{1, \ldots, N\}$ with having $\mathrm{K}(N)$ and $\lambda(N)$ both take on the last remaining value:

$$\kappa(N) = \lambda(N) \in \mathcal{N} \setminus \mathcal{N}_{N-1}$$

Note that κ is a bijection from $\{1, \ldots, N\}$ to $\mathcal{N}$.

Third, consider the incentives of agents in the proposed new SPNE corresponding with tax profile $(\tau_i')_{i\in\mathcal{N}}$. We show by induction that every agent $i \in \mathcal{N}$ wants to propose the transfer functions g_{ijt}' $\forall j \in \mathcal{N}$ and engage in the same action profile a_i^* as in the old SPNE. We start with agent $\kappa(1)$. Define $\forall i \in \mathcal{N}$:

$$V_i(s(h); (\tau_{i'})_{i'\in\mathcal{N}})$$

$$\equiv \begin{cases} \bar{u}_i, & \exists i', j, t, t' \colon c_{i'jt} = c_{ji't'} = 1, s_{i'jt}^g(h) + s_{ji't'}^g(h) \neq 0 \\ v_i\left(s^a(h), \sum_{j\in\mathcal{N}} \sum_{t=1}^{T} c_{jit} s_{jit}^g(h) - \tau_i\big(s^a(h)\big)\right), & otherwise \end{cases}$$

One can rewrite the definition of an SPNE strategy profile as a strategy s^* such that $\forall i \in \mathcal{N}$ $\forall h \in H$:

$$V_i(s^*(h), (\tau_{i'})_{i'\in\mathcal{N}}) \geq V_i\left(\big(s_i(h), s_{-i}^*(h)\big), (\tau_{i'})_{i'\in\mathcal{N}}\right) \ \forall s_i \in S_i(h)$$

Inductively define for $n = 1, \ldots, N-1$ the values of transfer adjustments:

$$\Delta g_{\lambda(1)\kappa(1)} \equiv \tau_{\mathrm{K}(1)} - \tau_{\kappa(1)}'$$

$$\Delta g_{\kappa(1)\lambda(1)} \equiv -\Delta g_{\lambda(1)\kappa(1)}$$

$$\Delta g_{\lambda(n)\kappa(n)} \equiv \tau_{\mathrm{K(n)}} - \tau_{\kappa(n)}' + \sum_{i\in\mathcal{N}:\lambda(\kappa^{-1}(i))=\kappa(n)} \Delta g_{\kappa(n)i}$$

$$\Delta g_{\kappa(n)\lambda(n)} \equiv -\Delta g_{\lambda(n)\kappa(n)}$$

This definition is inductive because for any value of n for which all values $\Delta g_{\kappa(n)\lambda(n)}$ and $\Delta g_{\lambda(n)\kappa(n)}$ are already defined, $\Delta g_{\kappa(n)i}$ is already defined $\forall i\colon \lambda\big(\kappa^{-1}(i)\big) = \kappa(n)$.[23] Define all other values of Δg_{ij} as the zero function. Note that $\forall n = 1, \dots, N-1$:

$$\sum_{i \in \mathcal{N}} \Delta g_{i\kappa(n)} = \Delta g_{\lambda(n)\kappa(n)} - \sum_{j \in \mathcal{N}: \lambda(\kappa^{-1}(j)) = \kappa(n)} \Delta g_{\kappa(n)j} = \tau_i - \tau_i'$$

For completeness, noting that $\Delta g_{i\kappa(N)} = 0\ \forall i$:[24]

$$\sum_{j \in \mathcal{N}} \Delta g_{\kappa(N)j} = \sum_{i \neq \kappa(N)} [\tau_i' - \tau_i]$$

Letting $\Delta g_i \equiv \left(\Delta g_{ij}\right)_{j \in \mathcal{N}}$ denote the profile of transfer adjustments, we can then construct the strategy profile of the proposed new SPNE:[25]

$$\Delta\tau^{g}_{s_{it}}\left(\left(\left((a_{i't'}, g_{i't'})\right)_{t'=1}^{T}\right)_{i' \in \mathcal{N}}\right) \equiv s^{g}_{it}\left(\left(\left((a_{i't'}, g_{i't'} - \Delta g_{i'})\right)_{t'=1}^{T}\right)_{i' \in \mathcal{N}}\right) + \Delta g_i$$

$$\Delta\tau^{a}_{s_{it}}\left(\left(\left((a_{i't'}, g_{i't'})\right)_{t'=1}^{T}\right)_{i' \in \mathcal{N}}\right) \equiv s^{a}_{it}\left(\left(\left((a_{i't'}, g_{i't'} - \Delta g_{i'})\right)_{t'=1}^{T}\right)_{i' \in \mathcal{N}}\right)$$

$$\Delta\tau_{s_{it}}(h) \equiv \left(\Delta\tau^{a}_{s_{it}}(h), \Delta\tau^{g}_{s_{it}}(h)\right)$$

$$\Delta\tau_{s_i}(h) \equiv \left(\Delta\tau_{s_{it}}(h)\right)_{t=1}^{T}, \Delta\tau_{\mathrm{s}}(h) \equiv \left(\Delta\tau_{s_i}\right)_{i \in \mathcal{N}}, \Delta\tau^{a}_{s} \equiv \left(\left(\Delta\tau^{a}_{s_{it}}(h)\right)_{t=1}^{T}\right)_{i \in \mathcal{N}}$$

---

[23] For $n' = 1, \dots, n-1$ such that $\lambda(n') = \kappa(n)$, for ay $i = \kappa(n')$, $\Delta g_{\lambda(n')\kappa(n')}$ and $\Delta g_{\kappa(n')\lambda(n')}$ are already defined.

[24] We can prove this claim via induction for any such model. The claim is obvious when $N = 2$. Suppose it always holds when $|\mathcal{N}| = N - 1$ for $N > 2$. Then one can define $\Delta g_{\widetilde{\kappa(N-1)J}} = \Delta g_{\kappa(N)j}$, $\Delta g_{\widetilde{J\kappa(N-1)}} = -\Delta g_{\widetilde{\kappa(N-1)J}}$, $g_{\kappa(N)j} = 0$, and $\Delta \widetilde{g_{J\kappa(N)}} = -\Delta g_{\kappa(N)j}\ \forall j \neq \mathrm{K}(N-1)\colon \lambda\big(\kappa^{-1}(j)\big) = \kappa(N)$, with $\Delta g_{\widetilde{\kappa(\mathrm{N}),\kappa(\mathrm{N}-1)}} = \tau_{\kappa(N-1)} - \tau'_{\kappa(N-1)} + \sum_{j \in \mathcal{N} \setminus \{\kappa(N)\}} \Delta g_{\widetilde{\kappa(N-1)J}}$, $\Delta g_{\widetilde{\kappa(\mathrm{N}-1),\kappa(\mathrm{N})}} = -\Delta g_{\widetilde{\kappa(\mathrm{N}),\kappa(\mathrm{N}-1)}}$ and otherwise $\Delta \widetilde{g_{IJ}} = \Delta g_{ij}$. This system of changes in transfer functions, excluding those for agent $\mathrm{K}(N)$, corresponds with $\widetilde{\mathcal{N}} \equiv \mathcal{N} \setminus \{\kappa(N)\}$ with all agents connected. By inductive hypothesis and definition:

$$\sum_{j \in \mathcal{N} \setminus \{\kappa(N)\}} \Delta g_{\widetilde{\kappa(N-1)J}} = \sum_{i \neq \kappa(N-1), \kappa(N)} [\tau_i' - \tau_i]$$

Then by definition $\sum_{j \in \mathcal{N}} \Delta g_{\kappa(N)j} = \sum_{j \in \mathcal{N}} \Delta \widetilde{g_{\kappa(N)J}} = \Delta g_{\widetilde{\kappa(N)\kappa(N-1)}} = \sum_{i \neq, \kappa(N)} [\tau_i' - \tau_i]$.

[25] Analogously, $\Delta\tau_{s_{it}}(\emptyset) \equiv \left(s^{a}_{it}(\emptyset), s^{g}_{it}(\emptyset)\right)$ and so on.

In words, agents adjust their payments and otherwise respond as if other agents had not adjusted their payments accordingly. It is convenient to denote the mapping from histories to altered histories as if the payments were not adjusted as:

$$\Delta\tau_H\left(\left(\left((a_{it}, g_{it})\right)_{t=1}^{T}\right)_{i\in\mathcal{N}}\right) \equiv \left(\left((a_{it}, g_{it} - \Delta g_i)\right)_{t=1}^{T}\right)_{i\in\mathcal{N}}$$

We now claim that $\Delta\tau_{s^*}$ is an SPNE for $(\tau_i')_{i\in\mathcal{N}}$ and for any SPNE s^* of tax code $(\tau_i)_{i\in\mathcal{N}}$ when $\forall a \in A$:

$$\sum_{i\in\mathcal{N}} \tau_i(a) = \sum_{i\in\mathcal{N}} \tau_i'(a)$$

This is because the above condition implies that $\forall h, i$:

$$V_i\left(\Delta\tau_s(h), \left(\tau_{i'}'\right)_{i'\in\mathcal{N}}\right) = V_i\big(s\big(\Delta\tau_H(h)\big), (\tau_{i'})_{i'\in\mathcal{N}}\big)$$

This claim requires proof, but upon doing so, it's clear that in every history every agent best responds to all other agents by choosing $\Delta\tau_{s_i^*}(h)$ because they were best responding under the old tax code with some history $\Delta\tau_H(h)$ and equilibrium s^*.

Pick $i \in \mathcal{N}$ and $h \in H$. The claim is true when $\exists i', j, t\colon c_{i'jt} = 1, s_{i'jt}^{g}\big(\Delta\tau_H(h)\big) + s_{ji't}^{g}\big(\Delta\tau_H(h)\big) \neq 0$, since then for the same values of i', j, and t:

$$\Delta\tau_{s_{i'jt}}^{g}(h) + \Delta\tau_{s_{ji't}}^{g}(h) = s_{i'jt}^{g}\big(\Delta\tau_H(h)\big) + \Delta g_{i'j} + s_{ji't}^{g}\big(\Delta\tau_H(h)\big) + \Delta g_{ji'} \neq 0$$

And so:

$$V_i\left(\Delta\tau_s(h), \left(\tau_{i'}'\right)_{i'\in\mathcal{N}}\right) = \bar{u}_i = V_i\big(s\big(\Delta\tau_H(h)\big), (\tau_{i'})_{i'\in\mathcal{N}}\big)$$

Suppose otherwise, i.e.:

$$\forall i', j, t, c_{i'jt} = 1 \Rightarrow s_{i'jt}^{g}\big(\Delta\tau_H(h)\big) + s_{ji't}^{g}\big(\Delta\tau_H(h)\big) = 0$$

This means that $V_i\big(s\big(\Delta\tau_H(h)\big), (\tau_{i'})_{i'\in\mathcal{N}}\big) = \mathcal{v}_i\big(s^a\big(\Delta\tau_H(h)\big), \sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit} s_{jit}^g\big(\Delta\tau_H(h)\big) - \tau_i\big(s^a(j)\big)\big)$, and also that:

$$\forall i', j, t, c_{i'jt} = 1 \Rightarrow \Delta\tau_{s_{i'j}}^g(h) + \Delta\tau_{s_{ji'}}^g(h) = 0$$

Which implies that $V_i\big(\Delta\tau_{s_i}, (\tau_{i'})_{i'\in\mathcal{N}}\big) = \mathcal{v}_i\left(\Delta\tau_s^a(h), \sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit}\Delta\tau_{s_{jit}}^g(h) - \tau_i'\big(s^a(h)\big)\right)$.

Additionally:

$$\Delta\tau_s^a(h) = s^a\big(\Delta\tau_H(h)\big)$$

Finally, we want to show that:

$$\sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit}\Delta\tau_{s_{jit}}^g(h) - \tau_i'\big(s^a(h)\big) = \sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit} s_{jit}^g\big(\Delta\tau_H(h)\big) - \tau_i\big(s^a(h)\big)$$

We show this by demonstrating that:

$$\sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit}\left[\Delta\tau_{s_{jit}}^g(h) - s_{jit}^g\big(\Delta\tau_H(h)\big)\right] = \sum_j \Delta g_{ji} = \tau_i - \tau_i'$$

The left-hand equality comes from the definition of $\Delta\tau_{s_{jit}}^g(h)$. The right-hand equality is obviously true for any $i \neq \kappa(N)$.[26] It is true for $i = \kappa(N)$ because:

$$\sum_{i'\in\mathcal{N}} \tau_{i'} = \sum_{i'\in\mathcal{N}} \tau_{i'}' \Rightarrow \tau_i - \tau_i' = -\sum_{i'\neq i}[\tau_{i'}' - \tau_{i'}] = -\sum_{j\in\mathcal{N}} \Delta g_{\kappa(N)j} = \sum_{i'} \Delta g_{i'i}$$

Thus:

[26] For $i \neq \kappa(N)$, $\sum_j \Delta g_{ji} = \Delta g_{\lambda\left(\kappa^{-1}(i)\right)i} - \sum_{j:\lambda(\kappa^{-1}(j))=i} \Delta g_{ij} = \tau_i - \tau_i^{'}$.

$$V_i\left(\Delta\tau_{s_i}, (\tau_{i'})_{i'\in\mathcal{N}}\right) = v_i\left(\Delta\tau_s^a(h), \sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit}\Delta\tau_{s_{jit}}^g(h) - \tau_i'\left(s^a(h)\right)\right)$$

$$= v_i\left(s^a\left(\Delta\tau_H(h)\right), \sum_{j\in\mathcal{N}}\sum_{t=1}^{T} c_{jit}s_{jit}^g\left(\Delta\tau_H(h)\right) - \tau_i\left(s^a(h)\right)\right)$$

$$= V_i\left(s\left(\Delta\tau_H(h)\right), (\tau_{i'})_{i'\in\mathcal{N}}\right)$$

Therefore, $V_i\left(\Delta\tau_s(h), \left(\tau_{i'}'\right)_{i'\in\mathcal{N}}\right) = V_i\left(s\left(\Delta\tau_H(h)\right), (\tau_{i'})_{i'\in\mathcal{N}}\right)\ \forall h, i$, implying that $\Delta\tau_{s^*}$ is an SPNE for tax code $(\tau_i')_{i\in\mathcal{N}}$. ■

Chapter 3

Doctors Need Sleep: A Theory of Optimal Production

by Benjamin Glass

A Walrasian equilibrium with production cannot exist when firms have unboundedly increasing returns to scale technology. If any price-taking firm chooses to produce at all, then it could achieve infinite profit by producing an infinite amount. Thus, the difficulty of increasing returns to scale in perfect competition results from firms acting without accounting for input supply constraints.

1. Introduction

There is no such thing as decreasing returns to scale production technology. Any productive process performed once can be replicated, so that doubling inputs can easily double outputs. In fact, if any resources are spent discovering improvements in productive techniques, then the same argument implies that returns to scale are necessarily increasing. This paper offers the first general model of price-taking agents with increasing returns to scale production technology as a competitive benchmark for efficiency.

The argument against decreasing returns to scale goes back to Koopmans (1959), yet the assumption persists. This may be due to the usefulness of decreasing returns to scale in explaining such basic economic phenomena as supply curves and profit taxes. As Kimball (2017) notes, there are multiple justifications for the assumption. Empirically, we cannot always observe input quality, so it may appear to us that firms face decreasing returns to scale as they use more lower quality inputs.   Theoretically, many standard results go through with slight modification if we suppose that there are some factors that firms cannot trade between each other, even as these factors may be freely used within the firms to which they are attached. For instance, if each firm has non-transferable managerial oversight that does not appear in the production function, then the welfare theorems should go through so long as we understand that we are referring to constrained Pareto efficiency subject to the immobility of factors between firms.

Yet these justifications reach their limit when considering the intersection of competitive production and public policy. The problems are two-fold. One, even if we imagine information asymmetries or contractibility problems such that factors cannot transfer between firms, surely the supply of these factors tends to change with market conditions. For instance, if the corporate income tax declines, one would expect managers to attempt to realize higher profits by supplying more managerial oversight. Two, and perhaps more critically, the assumption of the immobility of factors between firms has a very natural policy implication: the government should force firms to merge together. By merging together two firms with fixed factors, the government creates a single entity that in theory could operate as the two distinct firms previously had. Thus, there is no risk to productivity, but the possibility remains that the newly formed firm could benefit by moving around previously fixed factors.

Regarding the theoretical recommendation to forcibly merge firms, there are three counterarguments. The first asks why these firms did not previously willingly merge. I suggest that the information asymmetry or contractibility issues that keep factors fixed likely also thwart firms from mutually agreeing to merge. In any case, if it is not theoretically possible for the merged firm to be less productive, the failure of these firms to merge on their own does not introduce this possibility. Second, one may suggest that the merged firm is more difficult to manage. But this counterargument already supposes that the merged firm has lost some of its managerial labor. Instead, the merged firm is free to ignore the merger and act like two firms. And third, merging all firms together would create an entity so large that one would expect it to understand how its decisions affected prices. Thus, there may be limits to how much we can merge firms together without creating distortions. For instance, if firms are all single-product and compete in prices, then we could imagine Bertrand competition yielding competitive prices so long as at least two firms remain in each market.[1] In any case, so long as firms are small price-takers, the government can increase efficiency by forcing a small number of them to merge.

In response to these theoretical difficulties, I offer a model that does away with firms entirely. Instead, agents are directly endowed with production technology. I impose no convexity assumption on this technology. Instead, I allow for a set of goods that agents know they can supply for production but cannot purchase thereafter. In this way, agents account for some feasibility constraints in the economy. In doing so, they produce efficiently. I show in this model that a price-taking equilibrium is always efficient. Furthermore, any Pareto efficient allocation is supported by a price-taking equilibrium with transfers.

For instance, imagine potential doctors deciding how to allocate their time. On the one hand, there is a wage rate for their labor. But another opportunity is to sell some time and

[1] Such a result would hold if the merged firms had constant returns to scale up to a very large quantity.

other resources for training to make their labor able to produce valuable medical services. Potential doctors with a greater desire or capacity to practice medicine will sacrifice the necessary resources to obtain medical training. In doing so, they account for the long hours they will choose to work afterwards, explicitly aware that their increased earning will not permit them to repurchase their leisure on the open market. Upon achieving medical degrees, doctors may feel strong incentives to supply most of their time. Yet most doctors, like most people, require some amount of sleep. Even if they don't, doctors only have so much time in the day. Thus, even if they take all prices as given, their supply of medical care remains limited.

Critical to this story is the fundamentally limited size of medical practices. This model excludes cases of natural monopolies and public goods. For instance, if a radio broadcast can be produced for free upon the payment of a fixed cost in terms of resources, then no price-taking equilibrium can sustain the provision of public radio.[2] Instead, if there is a continuum of doctors, then the feasible set of allocations between leisure and the provision of medical care is convex. I model agents as in a continuum to ensure convexity of the set of feasible bundles.

The notion of convexity arising from a non-atomic measure of agents extends at least to Aumann (1966). Since the Second Welfare Theorem fails to hold in atomic economies with increasing returns to scale due to the non-convexity in the set of feasible bundles, one might hope that having a continuum of firms would solve this problem. This insight partially holds. Any model of a natural monopoly or of public goods inherently refers to atomic producers. For instance, if one must build a radio tower to produce radio broadcasts, then an appropriate

[2] The exception is if some inputs have negative market value, such as advertising content. Of course, advertising is inconsistent with a competitive economy of price-taking agents.

model requires a positive measure of inputs such as labor and steel to produce any amount of a radio broadcast. Still, so long as we avoid fixed cost production on a large scale, one can model production with a continuum of firms and obtain a convex set of feasible outcomes. See Hildenbrand (1969, 1974).

It may appear strange that a continuum of agents yields the welfare theorems, since any firm in the continuum still cannot achieve non-negative profit while producing with increasing returns to scale technology in equilibrium. The resolution of this apparent paradox is that a continuum of increasing returns to scale firms generally yields a set of feasible allocations that is not compact, and so may not have a Pareto efficient allocation. Without any Pareto efficient allocations, there are no equilibria. Intuitively, one can achieve arbitrarily more efficient production by having an arbitrarily small measure of firms scale up production. Yet this process of increasing efficiency cannot result in the most efficient result of a single firm producing, since that firm cannot yield a positive measure of production. Instead, the process of increasing efficiency yields a sequence of feasible allocations that approaches a boundary point not contained in the set of feasible bundles.

In contrast, the model I propose does yield a compact set of feasible bundles. If agents have continuous preferences, it follows that there are Pareto efficient outcomes. The Second Welfare Theorem then implies that one can obtain those outcomes with an appropriate lump-sum transfer scheme.

The contribution of this paper is primarily normative. The Second Welfare Theorem says that any Pareto efficient allocation can be realized by a decentralized price mechanism with transfers. This allows the planner to obtain any desired efficient allocation without relying on knowledge privately held by agents.[3] The paper also modestly contributes positively to our understanding of why apparently more efficient resource allocations arise in more competitive market settings. Walrasian equilibrium cannot explain how efficient allocations may occur in a price-taking environment with unboundedly increasing returns to scale production technology, as no such equilibrium can exist. In contrast, I show that the non-tradability of inputs allows for the existence of a Pareto efficient allocation, to which there corresponds a price-taking equilibrium with transfers.

This paper does not present the first model deviating from firms maximizing profits subject to productive technology. Models such as Grossman and Hart (1981, 1983) formalize information constraints preventing efficient production. But these considerations do not preserve the argument for decreasing returns to scale. Perhaps principle-agent problems make each worker employed and unit of capital rented less productive than is theoretically possible. All the same, whatever productive process one can employ subject to both feasibility and incentive-compatibility constraints with a given combination of factor inputs, one can surely repeat this process a second time to get at least twice the outputs with twice the inputs. Thus, Koopmans's (1959) argument against decreasing returns to scale survives treating incentive compatibility constraints like technological constraints.

There is substantial literature on notions of general equilibrium with increasing returns to scale. One literature on monopolistic competition attempts to model inefficient equilibria that arise form firms exercising market power. Another literature considers regulated

[3] However, that equilibrium with transfers is generally not unique suggests the planner may have difficulty assuring that a transfer scheme will assuredly yield the desired efficient allocation.

monopolies subject to pricing rules.[4] In contrast, I am aware of two papers that describe decentralized equilibria that both allow for increasing returns to scale production technology and at least sometimes obtains efficient equilibria.[5] The paper perhaps most related to this one is Dehez and Drèze (1988). They have firms set prices competitively and take as given constraints on how much quantity they can supply. Thus, they also have agents taking account of supply constraints. Yet their model does not necessarily yield efficient equilibria when production technology is not convex. Edlin et al. (1998) consider an otherwise competitive economy except for a perfectly price-discriminating monopolist. Their notion of equilibrium is typically efficient, but some Pareto efficient allocations cannot be sustained even with transfers.[6]

There is also a literature on incomplete financial markets in general equilibrium. My model corresponds with the notion of entrepreneurial equilibrium described in Magill and Quinzii (1996) when interpreting goods as consumption in various states. I am unaware of any previous literature discussing the (constrained) efficiency of a competitive economy of sole proprietorships with incomplete markets.

One may also conceive of my model as one of non-linear pricing, where the "price" of non-tradable inputs is the marginal dollar value of the utility of that input as it is utilized in production. This would relate my work to Chavas and Briec (2012), who demonstrate that

---

[4] See Heal (1999) for a literature review.

[5] By "decentralized", I mean the owners of the means of production would choose the production schedule that results in equilibrium, taking as given the environment determined by the Walrasian auctioneer.

[6] Vazirani (2013) describes a model in which income is exogenously given, but one could reinterpret it as a small open economy in which agents must sell their endowments for income at international prices, and then use that income to buy consumption goods from a perfectly price-discriminating importer. In this case, the welfare theorems obtain. However, this model still does not allow for increasing returns to scale.

economic efficiency results from a non-linear price taking equilibrium even when the representative firm has non-convex production technology. But my model explicitly differs from theirs in offering a decentralized efficient equilibria.

The paper is organized as follows. Section 2 lays out the model. Section 3 defines Pareto efficiency and equilibrium. Section 4 contains the welfare theorem results. Section 5 demonstrates existence of Pareto optima. Section 6 discusses modeling details and plans for future work. Section 7 concludes. The proof of theorem 2 is relegated to an appendix.

2. Model

Let $\theta \in \Theta$ indicate consumer type, where Θ is finite. For each type θ, there is mass $\lambda_\theta > 0$ agents of that type. Let $l \in \{1, \dots, L\}$ index the finite number of goods. For each type θ, there is a continuum of agents $i \in [0,1]$ with the same endowment vector $\omega_\theta \equiv (\omega_{\theta l}) \in \mathbb{R}_+^L$ and production technology $Y_\theta \subseteq \mathbb{R}^L$. Yet some inputs $\mathcal{I} \subset \{1, \dots, L\}$ cannot be purchased on the market.[7] Instead, they can only be used as productive inputs or consumed. Thus, an agent of type θ facing price vector p chooses:

$$(c, y) \in \mathbb{R}_+^L \times Y_\theta : p * c \leq p * [\omega_\theta + y], c_l = \omega_{\theta l} + y_l \; \forall l \in \mathcal{I}$$

Here, $c = (c_l)_{l=1}^L$ denotes a consumption vector, and for any agent of type θ, I denote by $c_{\theta l}$ and $\omega_{\theta l}$ that agent type's consumption and endowment respectively of good l. Define the set of feasible production vectors as:

$$\tilde{Y}_\theta \equiv \{y \in Y_\theta | \; y_l \in [-\omega_{\theta l}, 0] \; \forall l \in \mathcal{I}\}$$

I assume that non-purchasable inputs are essential to production. Formally:

[7] The analogue of the First Welfare Theorem implicitly assumes that $\mathcal{I} \neq \{1, \dots, L\}$.

Assumption 1: For any θ, $\tilde{Y}_\theta$ is compact and $y \in Y_\theta \Rightarrow y_l \leq 0 \; \forall l \in \mathcal{I}$.

This precludes production technology in which a fixed expenditure of inputs permits access to linear production technology using inputs outside of set $\mathcal{I}$.

All consumers of type θ maximize preferences $\succcurlyeq_\theta$ on $\mathbb{R}^L_+$. A result analogous to the Second Welfare Theorem will require that these preferences are continuous.

Assumption 2: For any θ, if $c' \succ_\theta c$, then $\exists \epsilon > 0$ such that $c'' \succcurlyeq_\theta \tilde{c} \; \forall c'' : \|c'' - c'\|, \|\tilde{c} - c\| < \epsilon$.

3. Equilibrium

To define equilibrium requires integrals of functions over the unit interval. To ease notation, define for any measurable set $E \subseteq [0,1]$:[8]

$$\mathbb{P}(E) \equiv \int_{i \in E} di$$

In words, $\mathbb{P}$ is the probability (via a uniform distribution) that a randomly selected element of $[0,1]$ is in E. A function $E(\theta)$ mapping from type to (measurable) subsets of the interval $[0,1]$ occurs *almost everywhere*, or for *almost every* (a.e.) agent, if:

$$\mathbb{P}\big(E(\theta)\big) = 1 \; \forall \theta$$

[8] This is a Lebesgue integral. The set of measurable sets E is the Borel sigma algebra on $[0,1]$.

A *feasible resource allocation* is a vector of pairs of a.e. continuous functions on the unit interval $(c_\theta, y_\theta)_{\theta\in\Theta}$ satisfying the feasibility constraints:[9]

1. $y_\theta(i) \in Y_\theta \;\forall i \;\forall \theta$
2. $c_{\theta l}(i) = \omega_{\theta l}(i) + \mathrm{y}_{\theta \mathrm{l}}(i) \;\forall l \in \mathcal{J} \;\forall i \;\forall \theta$
3. $\sum_\theta \lambda_\theta \int_0^1 c_{\theta l}(i)di = \sum_\theta \lambda_\theta \int_0^1 [\omega_\theta(i) + y_\theta(i)]di$

The first condition says that all production is feasible. The second condition says that agents cannot buy or sell inputs $\mathcal{J}$.[10] The third condition says that there is enough production to satisfy consumption. For convenience, we denote by $s_\theta(i) \equiv \omega_\theta + y_\theta(i)$ the vector of supply due to agent i of type θ.

A feasible resource allocation is *Pareto efficient* if no alternative feasible allocation both makes a.e. agent at least as well off and leaves some agents strictly better off.

Definition: A feasible resource allocation $(c_\theta, y_\theta)_\theta$ is <u>Pareto efficient</u> if for any feasible resource allocation $(c'_\theta, y'_\theta)_\theta$:

$$\forall \theta \; \mathbb{P}(c'_\theta \succcurlyeq_\theta c_\theta) = 1 \Rightarrow \forall \theta \; \mathbb{P}(c'_\theta \sim_\theta c_\theta) = 1$$

An equilibrium in this economy has a.e. agents choosing input supply, production, and consumption to maximize preferences subject to a price vector and technology constraints, as

[9] I imposed continuity a.e. of c_θ and y_θ, rather than the weaker assumption of integrability, to avoid issues with the existence of non-measurable sets of agents who disagree about which allocation they prefer.
[10] If one reinterprets goods as consumption in various states in an economy with incomplete financial markets, then including this condition suggests a description of *constrained* feasible allocations. Similarly, the subsequent description of Pareto efficiency would refer to constrained efficiency.

well as exogenous government transfers. In addition, the choices of these agents clear all markets. Formally, for any $T \in \mathbb{R}$, define the budget set of an agent of type θ as:

$$B_\theta(p, \omega, y, T) \equiv \{c \in \mathbb{R}_+^L \,|\, p * c \le p * [\omega + y] + T, c_l = \omega_{\theta l} + y_l \; \forall l \in \mathcal{I}\}$$

The government exogenously imposes the integrable transfer functions $T_\theta: [0,1] \to \mathbb{R}$. We say these transfers satisfy the balanced budget constraint if $\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$. It will be convenient to write the set of choices of consumption bundles available to the consumer as:

$$B_\theta(p, \omega, T) \equiv \{c \in \mathbb{R}_+^L \,|\, \exists y \in Y_\theta : c \in B_\theta(p, \omega, y, T)\}$$

Definition: An equilibrium $((c_\theta, y_\theta)_{\theta \in \Theta}, p)$ is made up of integrable functions and satisfies:

1. $c_\theta(i) \in B_\theta\big(p, \omega_\theta, T_\theta(i)\big) \; \forall i \; \forall \theta$
2. A.e. $\forall c' \in B_\theta\big(p, \omega_\theta, T_\theta(i)\big)$:

$$c_\theta(i) \succcurlyeq_\theta c'$$

3. $\sum_\theta \lambda_\theta \int_0^1 c_{\theta l}(i)di = \sum_\theta \lambda_\theta \int_0^1 s_\theta(i)di$

The first two conditions say that consumers maximize preferences subject to constraints. The third condition states that all markets clear.

4. The Welfare Theorems

We are now ready to demonstrate that the analogues of the welfare theorems hold in this new model, beginning with the analogue of the First Welfare Theorem. The standard result uses an assumption that all agents have locally non-satiated preferences. However, now agents are aware of the supply constraints of certain goods, so that certain forms of locally non-satiated preferences may still yield inefficient equilibria.

For instance, suppose $\Theta = \{1,2\}$ and $L = 2$, where agents of type 1 have preferences represented by $u_1(c_{11}, c_{12}) = \min\{c_{11}, c_{12}\}$, while agents of type 2 have preferences represented by $u_2(c_{21}, c_{22}) = c_{22}$. Let $\omega_1 = (1,0)$, $\omega_2 = \mathbf{0}$, and:

$$Y_1 = \{(y_1, y_2) | 0 \leq y_2 \leq \mathbb{I}(\mathrm{y}_1 \leq -1)\}$$

$$Y_2 = \{\mathbf{0}\}$$

Here $\mathbb{I}$ denote the indicator function mapping true statements to one and false statements to zero, while $\mathbf{0}$ denotes the zero vector.[11] In words, agent type one can produce one unit of good two only by giving up their endowment of good one.

It is Pareto inefficient for agents of the first type to consume their endowments, since they could be just as well off producing and giving their produced good to agents of type 2. Yet for a set $\mathcal{J} = \{1\}$ and a price vector $p = (0,1)$, agents of type 1 choose (c_{11}, c_{12}) such that either $c_{11} = 0$ or $c_{12} = 0$. They can only achieve utility of zero, and so can optimally choose $c_1 = (1,0)$ and $y_1 = \mathbf{0}$. Meanwhile, agents of type 2 can only choose $c_2 = \mathbf{0}$. Thus, it is an equilibrium resource allocation for all agents to consume their initial endowments, despite this outcome being inefficient. Critical to this reasoning is that agent type one cannot purchase good 1, as otherwise they would do better to choose $c_1 = (1,1)$, violating market-clearing.

To ensure equilibrium efficiency, we impose a stronger assumption than local non-satiation. Informally, it is local non-satiation with respect to tradable goods. This condition remains weaker than assuming strictly monotonic preferences.

First Welfare Theorem: Any equilibrium is Pareto efficient if for all θ and all $c \in \mathbb{R}^L$:

---

[11] I will generally use $\mathbf{0}$ to denote the zero vector for an arbitrary number of goods L.

$$\forall \epsilon > 0\ \exists c' = (c_l')_{l=1}^{L}: c' \succ_\theta c, c_l' = c_l\ \forall l \in \mathcal{J}, \|c' - c\| < \epsilon$$

In words, an arbitrarily small change in consumption of non-tradable goods can make the consumer better off.

Proof: This proof is completely standard. First, note that the assumption on preferences implies that agents spend all their income:

$$p * c_\theta(i) = p * s_\theta(i) + T_\theta(i)\ \forall i\ \forall \theta$$

Then feasibility shows that $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium only if the transfers satisfy the balanced budget condition:

$$\sum_\theta \lambda_\theta \int_0^1 [p * s_\theta(i) + T_\theta(i)]di = \sum_\theta \lambda_\theta \int_0^1 p * c_\theta(i)di = \sum_\theta \lambda_\theta \int_0^1 p * s_\theta(i)di$$

$$\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$$

Suppose that there were feasible resource allocations $(c_\theta, y_\theta)_\theta$ and $(c_\theta', y_\theta')_\theta$ such that:

$$\forall \theta\ \mathbb{P}\big(c_\theta'(i) \succcurlyeq_\theta c_\theta(i)\big) = 1$$

Assume that $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium for some price vector $p \in \mathbb{R}^L$. The assumption on preferences implies that $p * [(c_\theta' - s_\theta') - (c_\theta - s_\theta)]$ is an integrable function a.e. non-negative. Yet market-clearing and feasibility implies that:

$$\sum_\theta \lambda_\theta \int_0^1 p * \big[\big(c_\theta'(i) - s_\theta'(i)\big) - \big(c_\theta(i) - s_\theta(i)\big)\big]di = -\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$$

Thus, we know that $p * \big(c_\theta'(i) - s_\theta'(i)\big) = p * \big(c_\theta(i) - s_\theta(i)\big)$ a.e., implying that a.e.:

$$c_\theta'(i) \in \underset{B_\theta(p,\omega_\theta,T_\theta(i))}{\operatorname{argmax}} \succcurlyeq_\theta$$

By symmetry, I can conclude that:

$$\forall\theta\ \mathbb{P}(c'_\theta \sim_\theta c_\theta) = 1$$

And so $(c_\theta, y_\theta)_\theta$ is Pareto efficient. ■

In the proof of the analogue to the First Welfare Theorem, an equilibrium is efficient because the existence of any feasible Pareto improvement would induce a positive measure of agents to deviate from the proposed equilibrium. However, our main concern is the ability to sustain an equilibrium at an arbitrary Pareto efficient allocation using the price mechanism. The Second Welfare Theorem traditionally argues for this possibility under assumptions of convex production technology. The idea behind this result is to impose assumptions such that the set of feasible bundles is convex. Then the price vector yielding equilibrium is the supporting hyperplane on the Pareto efficient allocation. However, if firms have increasing returns to scale production technology, then the set of feasible allocations is not convex, and so there may not exist a supporting hyperplane.

Our current setup resolves this difficulty by assuming a non-atomic measure of agents. Each of these agents has increasing returns to scale production technology, but their supply constraints bound their sets of possible production vectors. Thus, the feasible production set remains convex.

Second Welfare Theorem: Under assumptions (1) and (2), any Pareto efficient allocation $(c_\theta, y_\theta)_\theta$ has a corresponding price vector p such that $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium with integrable transfers $T_\theta\colon [0,1] \to \mathbb{R}$ satisfying the balanced budget condition:

$$\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$$

The outline of the proof follows seven steps like the standard proof of the Second Welfare Theorem. First, let V denote the set of aggregated Pareto improvements:

$$V \equiv \left\{ \begin{array}{c} C \in \mathbb{R}^L \mid \exists c'_\theta \ \forall \theta: c'_\theta(i) \succcurlyeq_\theta c_\theta(i) \ \forall i \ \forall \theta, \exists \theta: \mathbb{P}\big(c'_\theta(i) \succ_\theta \ c_\theta(i)\big) > 0, C = \\ \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i) di \end{array} \right\}$$

One can show that V is an open convex set. Next, let R denote the set of aggregated feasible allocations:

$$R \equiv \left\{ X \in \mathbb{R}^L \mid \exists (\mathrm{c}'_\theta, y'_\theta)_\theta : \forall \theta \ \forall i \ c_{\theta l}(i) = \omega_{\theta l} + \mathrm{y}_{\theta l} \ \forall l \in \mathcal{J} \ \forall i, y'_\theta(i) \in Y_\theta, X = \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i) di \right.$$

$$\left. = \sum_\theta \lambda_\theta \int_0^1 [\omega_\theta + y'_\theta(i)] di \right\}$$

One can also show that R is convex. Third, if V is empty, then the set of agents who are not globally sated at allocation $(c_\theta, y_\theta)_\theta$ has measure zero, so $p = \mathbf{0}$ yields an equilibrium. We now assume otherwise, so that V and R are non-empty and disjoint. Then the Separating Hyperplane Theorem implies that:

$$\exists p \in \mathbb{R}^L, v \in \mathbb{R}: p * X \leq v \leq p * C \ \forall X \in R \ \forall C \in V$$

In words, there is a price vector p such that feasible allocations have value no greater than v, while Pareto improvements upon $(c_\theta, y_\theta)_\theta$ have value no less than v. In fact, by the openness of V:

$$v < p * C \ \forall C \in V$$

Fourth, still assuming $V \neq \emptyset$, one can show that the aggregation of the Pareto efficient allocation $(c_\theta, y_\theta)_\theta$ is on the boundary of the set V:

$$\sum_\theta \lambda_\theta \int_0^1 c_\theta(i) di \in cl(V)$$

Fifth, one can then note that the aggregation of allocation $(c_\theta, y_\theta)_\theta$ has market value v:

$$p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i) di = v$$

Sixth, define transfers $T_\theta(i) \equiv p * [c_\theta(i) - s_\theta(i)]\ \forall i\ \forall \theta$. Such a definition yields:

$$\sum_\theta \lambda_\theta \int_0^1 T_\theta(i) di = p * \sum_\theta \lambda_\theta \int_0^1 [c_\theta(i) - s_\theta(i)] di = 0$$

And seventh, we can confirm that $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium. We already know that:

$$c_\theta(i) \in B_\theta\big(p, \omega_\theta, T_\theta(i)\big)\ \forall i\ \forall \theta$$

The greater difficulty is to show that $\forall \theta$:

$$\mathbb{P}\left(c_\theta(i) \in \underset{B_\theta(p,\omega,T_\theta(i))}{\operatorname{argmax}} \succcurlyeq_\theta \right) = 1$$

To see this, suppose for the sake of contradiction that for some type θ^*:[12]

$$\mathbb{P}\left(c_{\theta^*}(i) \notin \underset{B_\theta(p,\omega,T_\theta(i))}{\operatorname{argmax}} \succcurlyeq_{\theta^*} \right) > 0$$

Then one can show that there exists some resource allocation $(c'_\theta, y'_\theta)_\theta$ such that for all agents i of all types θ:

$$c'_\theta(i) \succcurlyeq_\theta c_\theta(i)$$

Furthermore, there is a positive measure of agents of type θ^* who strictly prefer this new allocation:

$$\mathbb{P}\big(c'_{\theta^*}(i) \succ_{\theta^*} c_{\theta^*}(i) \big) > 0$$

---

[12] Really, suppose that the set $\left\{ c_{\theta^*}(i) \notin \underset{B_{\theta^*}(p,\omega,T_{\theta^*}(i))}{argmax} \succcurlyeq_{\theta^*} \right\}$ contains a set of positive measure. That this is so if $((c_\theta, y_\theta)_\theta, p)$ is not an equilibrium with transfers is left to the appendix.

And the resource allocation is consistent with individual constraints:

$$\exists y'_\theta(i): c'_\theta(i) \in B_\theta\big(p, \omega, y'_\theta, T_\theta(i)\big)\ \forall i\ \forall \theta$$

But then:

$$v < p * \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di \leq p * \sum_\theta \lambda_\theta \int_0^1 s'_\theta(i)di + \sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di \leq v$$

The first inequality comes from $\sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di \in V$, the second comes from the satisfaction of budget constraints, and the third comes from $\sum_\theta \lambda_\theta \int_0^1 s'_\theta(i)di \in R$ & $\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$. Thus, we have our contradiction, demonstrating that no positive measure of agents can benefit from deviating from the proposed equilibrium.[13] Finally, feasibility ensures that the markets clear:

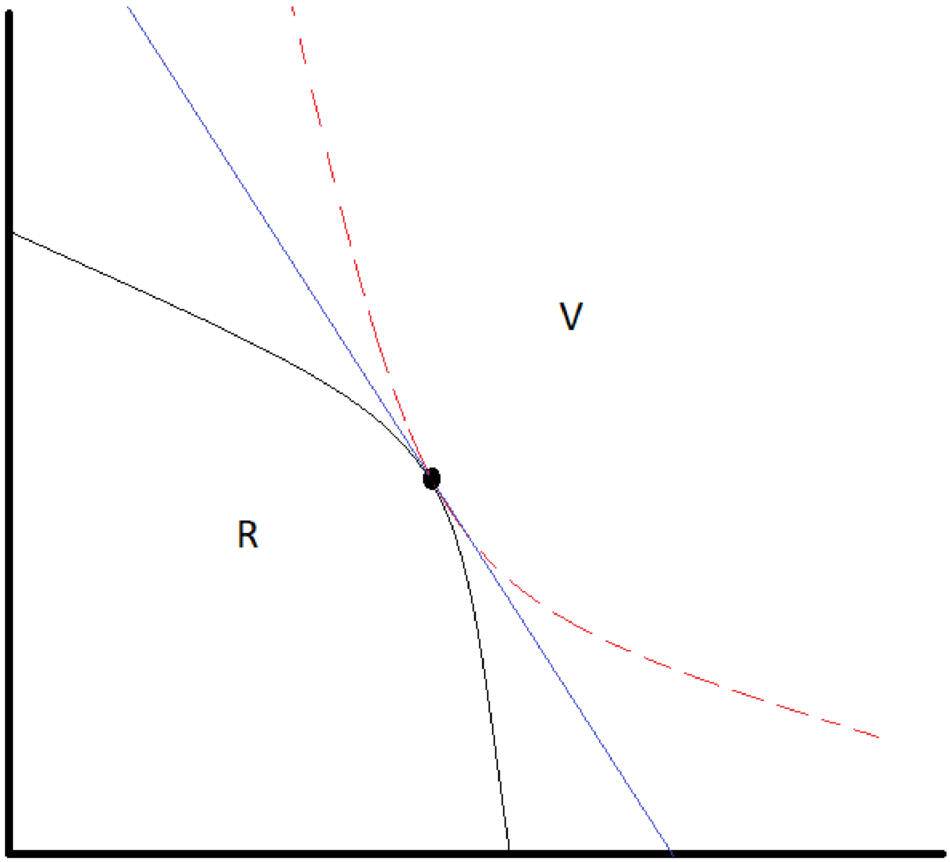
$$\sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di = \sum_\theta \lambda_\theta \int_0^1 s_\theta(i)di$$

We conclude that $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium with transfers $(T_\theta)_\theta$.

The above outline contains the main points of the proof. I relegate the details, including regarding the measurability and integrability of functions, to the appendix. Yet the proof outline is sufficient to note the similarity with standard demonstrations of the Second Welfare Theorem. The disjoint convexity of aggregations of feasible and Pareto improving sets, V and R respectively, yields a separating hyperplane that permits the construction of budget constraints for all agents.

[13] One can then adjust $((c_\theta, y_\theta, T_\theta)_\theta, p)$ by having the measure zero of agents not already optimizing with respect to their budget constraints to instead choose from $\underset{B_\theta(p,\omega,T_\theta(i))}{argmax} \succcurlyeq_\theta$.

Figure 3.1



The black curve delineates the boundary of the set R. The red dashed curve delineates the boundary of the (open) set V. The black dot represents the aggregation of the Pareto efficient allocation, on the boundary of R and V. This yields a hyperplane, denoted by the blue line, passing through the aggregated Pareto efficient allocation and separating the sets R & V.

5. Existence of Pareto Efficient Allocations

I have thus far performed demonstrations of results in the spirit of the classic First and Second Welfare Theorems. In summation, weak assumptions ensure that the resource allocations supported in equilibrium are exactly those which are Pareto efficient. The efficiency of equilibria comes from agents relying on prices to what degree they do not already internalize resource constraints. The ability to support efficient outcomes with equilibria arises from the convexity of sets V and R.

The convexity of both R and V is due to the non-atomic measure of agents. One can perform similar derivations using a continuum of firms as in the classic welfare theorems. But while a continuum of agents ensures that R is convex, it permits R to not be compact when production technology has increasing returns to scale. For instance, consider an economy with one type of agent and two goods. We suppress θ subscripts for convenience. The representative agent has $\omega_1 > 1$ and $\omega_2 = 0$. However, their preferences have utility representation $u(c) = c_2$. In words, the agent is endowed with one good, but desires the other good. The production technology is:

$$Y \equiv \{(y_1, \max\{|y_1 - 1|, 0\}) | y_1 \leq 0\}$$

In words production of one good is linear in the other good after payment of a fixed cost of one unit of the input good. If this were a description of a continuum of firms, production becomes increasingly efficient as a smaller measure of firms pay the fixed cost. However, any set of firms of measure zero cannot produce a positive measure of any good. Thus, the feasible values of good 2 are $[0, |\omega_1|)$. There is no Pareto efficient resource allocation.

But in the model of agent-endowed production technology, each agent must use their own endowed inputs to produce. Thus, the set of feasible values of good 2 is $[0, |\omega_1| - 1]$. One can confirm the existence of an equilibrium with price vector $p = (0,1)$.

More generally, the economy yields compact sets of feasible allocations under assumptions (1) and (2). This means Pareto efficient allocations exist, resolving the remaining difficulty due to increasing returns to scale production technology.

Theorem: Under assumptions (1) and (2), R is compact, and so Pareto efficient allocations exist.

Proof: R can be expressed as the Minkowski sum:

$$R = \sum_{\theta} \left[\{\omega_\theta\} + conv\left(\tilde{Y}_\theta\right) \right]$$

Since $\tilde{Y}_\theta$ is compact, so is $conv\left(\tilde{Y}_\theta\right)$. Finally, the Minkowski sum of compact convex sets is both compact and convex.

Let u_θ denote a continuous utility representation of preferences $\succcurlyeq_\theta$. From Hildenbrand (1974 pg. 62 theorem 3 and corollary), the set $U \equiv \left\{ \sum_\theta \lambda_\theta \int_0^1 u_\theta\left(c_\theta(i)\right) di \mid \sum_\theta \lambda_\theta \int_0^1 c_\theta(i) di \in R \right\}$ of feasible aggregate utility values is $conv(\{\sum_\theta \lambda_\theta u_\theta(c_\theta) \mid \sum_\theta \lambda_\theta c_\theta \in R\})$. Since R is compact, $\{\sum_\theta \lambda_\theta u_\theta(c_\theta) \mid \sum_\theta \lambda_\theta c_\theta \in R\}$ is compact, and so U is (convex and) compact.[14] ■

The first two theorems demonstrated the bijection between Pareto efficient allocations and equilibrium. This last theorem ensures that the use of a continuum of agents has not destroyed the existence of efficient allocations by erasing the boundary of R. Thus, the model of individual agent production with the supply limits of inputs $\mathcal{J}$ accounted for has resolved a theoretical difficulty with the welfare economics of increasing returns to scale.

6. Discussion and Extensions

I have shown that the welfare theorems and the existence of Pareto efficient allocations obtains in this model, in contrast with traditional models of increasing returns to scale production technology. This section addresses the most prominent outstanding issues with this model, particularly in contrast with more traditional competitive general equilibrium models.

[14] This means that any utility representations yield a closed interval of values for aggregate utility.

First, I have not imposed sufficient assumptions to ensure equilibrium existence for any transfer scheme $(T_\theta)_\theta$. Equilibrium existence even in an exchange economy requires assumptions not yet imposed. For instance, consider $\Theta = \{1,2\}$, $L = 2$, $Y_1 = Y_2 = \mathcal{J} = \emptyset$, with $u_1(c_1) = c_{12}$, $u_2(c_2) = c_{21} + c_{22}$, $\omega_1 = (1,0)$, $\omega_2 = \mathbf{0}$, and $T_1 = T_2 = 0$ with $\lambda_1 = \lambda_2$. In words, this is an endowment economy with both goods tradable, where one agent has a good that the other agent would benefit from consuming. One can confirm that this example satisfies assumptions (1) and (2).

For any price vector $p = (p_1, p_2)$:

$$p_1 > 0 \Rightarrow c_{11} = c_{21} = 0 \Rightarrow c_{11} + c_{21} = 0 < \omega_{11} + \omega_{21} = 1$$

Thus, any positive price for good 1 makes demand for good 1 less than supply. Yet:[15]

$$p_1 \leq 0 \Rightarrow c_{21} = \infty > \omega_{11} + \omega_{21} = 1$$

Therefore, the imposed assumptions remain insufficient for equilibrium existence even in an endowment economy.[16]

Also, there is no "profit" function, i.e. a function of the market values of inputs and outputs, on which one can impose a non-distortionary tax. For instance, consider $\Theta = \{1,2\}$, where $\omega_1 = (1,0)$, $u_1(c_1) = c_{11} + 2c_{12}$, $Y_1 = \{(y_{11}, |y_{11}|^2) |\, y_{11} \leq 0\}$, $\omega_2 = \mathbf{0}$, $u_2 = c_{22}$, and $Y_2 = \{\mathbf{0}\}$. In words, agent 1 has an input, say labor, that they can consume or transform into a consumption good. Agent 2 has nothing, cannot produce anything, and would like some of that consumption good. Finally, $\lambda_1 = \lambda_2 = 1$, so that there is a unit mass of agents of each type.

[15] Technically, $p_1 \leq 0$ precludes any solution to the preference-maximization problem.

[16] One could trivially extend this to have a non-tradable good with $L = 3$ and $\mathcal{J} = \{3\}$. In any case, one can check from Gale and Mas-Colell (1975) that the missing assumption for ensuring existence in an exchange economy is a condition ensuring positive income.

One can confirm that for any $\beta \in (0,1)$, $c_1 = (0, \beta)$ and $c_2 = (0, 1 - \beta)$ is efficient. Yet imposing a tax of $\tau > 0.5$ on output from the first agent, transferring the tax revenue to agent 2, and with a price vector $p = (0,1)$ yields agent 1's maximization problem:

$$\max_{c_{11} \in [0,1]} 2(1 - \tau)(1 - c_{11})^2 + c_{11}$$

This maximization problem yields corner solution $c_{11} = 1$ and $c_{12} = 0$. There is no tax revenue, and so $c_2 = \mathbf{0}$. Furthermore, note that in this equilibrium labor has no market value, and so there is no market value for inputs utilized. Thus, there is no tax on the market value of inputs and outputs that does not distort behavior.

More generally, endowing agents with production technology and having them optimize while accounting for supply constraints entirely rules out any construct of a profit-maximizing firm. Without these profit-maximizers, my model leaves no basis for considering the efficient boundaries of allocating resources within a managerial structure versus using the price mechanism. Future work should speak to the determination of when production happens by direction of planners versus via prices.

One approach is to let agents make agreements to work together on production, still such that they take supply constraints as given. For instance, consider $\Theta = \mathcal{J} = \{1,2\}$ with $\omega_1 = (1,0,0)$, $\omega_2 = (0,1,0)$, $\lambda_1 = \lambda_2 = 1$, and $Y = \{(y_1, y_2, \min\{|y_1|, |y_2|\}^2) | y_1, y_2 \leq 0\}$. Here the function Y refers to a technology jointly accessible by the two agent types. Both agents have preferences of the form $u_\theta(c_\theta) = c_{\theta 1} + c_{\theta 2} + 3c_{\theta 3}$. One can imagine a Walrasian auctioneer specifying that for any production plan chosen, agent type 1 gets fraction $\beta \in \left(\frac{1}{3}, \frac{2}{3}\right)$ of the resulting revenue, while agent type 2 gets fraction $1 - \beta$. An equilibrium would have both agents choose the same production plan, so that they happen to always agree. For instance, at $p = (1,1,1)$, both agents choose to produce one unit with $y_1 = y_2 = -1$, $c_1(i) = (0,0,\beta)\ \forall i$, $c_2(i) = (0,0,1-\beta)\ \forall i$, and so all markets clear. One can confirm that the resulting equilibrium is optimal.

This is only one example. It does not formally demonstrate the welfare theorems or the existence of Pareto efficient allocations. Furthermore, it appears such a model leaves undetermined the division of revenue between partners in an enterprise. Previous work on shareholder equilibria do not generally achieve even constrained notions of efficiency.[17] Future work should address these difficulties with alternative approaches to coordinating production in a competitive setting.

Finally, the production efficiency result from Diamond and Mirrlees (1971) describes optimal tax problems whose solutions do not permit aggregate production to reside in the interior of feasible aggregate consumption bundles. This result depends critically on the complete taxation of firm profits. Now there are no profits to tax. Thus, it remains an open question whether some form of the production efficiency result still obtains in this new setting. Still, the main result holds for centrally planned production.

I conjecture that any constrained optimal $(c_\theta, y_\theta)_\theta$ has $((y_{\theta l})_{l \notin \mathcal{J}})$ on the boundary of the feasible set conditional on $((y_{\theta l})_{l \in \mathcal{J}})_\theta$, given weak assumptions as in Diamond and Mirrlees (1971). I believe this holds because any agent i of type θ who has chosen $(y_{\theta l})_{l \in \mathcal{J}}$ chooses $(y_{\theta l})_{l \notin \mathcal{J}}$ to maximize profit. This is production efficiency if one interprets the consumption by

[17] Dreze (1974) created the canonical shareholder model with incomplete markets. Dierker, Dierker, and Grodal (2005) encapsulate previous results on the constrained inefficiency of the resulting equilibrium by showing that it does not satisfy even a minimal notion of constrained inefficiency with arbitrarily small income effects for consumers.

agents i and $i' \neq i$ of goods $l \in \mathcal{J}$ to be consumption of distinct goods. For instance, my consumption of leisure is distinct from your consumption of leisure.

7. Conclusion

When permitting increasing returns to scale production technology, traditional general equilibrium models could not simultaneously yield the Second Welfare Theorem and the existence of Pareto efficient allocations. But one can obtain both results by permitting agents to account for the limited supply of their own endowed inputs. This mirrors the real-world situation of doctors in private practice deciding how much labor to supply, understanding that the increased marginal productivity of their time after paying substantial fixed costs will not permit them to purchase their leisure on the market thereafter. Their desire for sleep and other leisure may limit their overall production level. Failing that, they only have so many hours in a day to supply.

These results have multiple substantial limits. First, they do not permit considerations of anything one might plausible call a firm. This precludes a substantial literature pertaining to optimal tax policy in general equilibrium. To what degree versions of these results survive in this new environment is an open question.

Second, the model does not allow for the consideration of public goods or natural monopolies. Any enterprise that requires a positive measure of inputs as a fixed cost to achieve production has no representation with a model of a non-atomic continuum of productive agents. There is a substantial literature on how one might regulate a natural monopoly or determine the optimal provision of a public good, but no result on how one might provide such goods efficiently in a decentralized environment. Of course, it is implausible that a natural monopoly would function as in a competitive setting. But even if one could make the sole provider of a good with fixed costs of production take prices as given, it is unclear whether there is a decentralized competitive environment which provides such goods efficiently.

References

Aumann, Robert J. "Existence of Competitive Equilibria in Markets with a Continuum of Traders." *Econometrica: Journal of the Econometric Society* (1966): 1-17.

Chavas, Jean-Paul, and Walter Briec. "On Economic Efficiency under Non-Convexity." *Economic Theory* 50.3 (2012): 671-701.

Coase, Ronald H. "The Nature of the Firm." *Economica* 4.16 (1937): 386-405.

Dehez, Pierre, and Jacques Dreze. "Competitive Equilibria with Quantity-Taking Producers and Increasing Returns to Scale." *Journal of Mathematical Economics* 17.2-3 (1988): 209 230.

Diamond, Peter A., and James A. Mirrlees. "Optimal Taxation and Public Production I: Production Efficiency." *The American Economic Review* 61.1 (1971): 8-27.

Dierker, Egbert, Hildegard Dierker, and Birgit Grodal. "Are Incomplete Markets Able to Achieve Minimal Efficiency?." *Economic Theory* 25.1 (2005): 75-87.

Dreze, Jacques H. "Investment under Private Ownership: Optimality, Equilibrium and Stability." *Allocation under Uncertainty: Equilibrium and Optimality*. Palgrave Macmillan, London, 1974. 129-166.

Edlin, Aaron S., Mario Ephemerella, and Walter P. Keller. "Is Perfect Price Discrimination Really Efficient?: Welfare and Existence in General Equilibrium." *Econometrica* (1998): 897 922.

Fisher, Irving. "The Theory of Interest, as Determined by Impatience to Spend Income and Opportunity to Invest it." *New York* (1930).

Gale, David, and Andreu Mas-Colell. "An Equilibrium Existence Theorem for a General Model Without Ordered Preferences." *Journal of Mathematical Economics* 2.1 (1975): 9-15.

Heal, Geoffrey M., ed. *The Economics of Increasing Returns*. Edward Elgar, 1999.

Hildenbrand, Werner. "Pareto Optimality for a Measure Space of Economic Agents." *International Economic Review* 10.3 (1969): 363-372.

Hildenbrand, Werner. Core and Equilibria of a Large Economy. Princeton, New Jersey: Princeton University Press, 1974.

Kimball, Miles. "There Is No Such Thing as Decreasing Returns to Scale." May 30th, 2017, blog.supplysideliberal.com/post/2017/5/29/there-is-no-such-thing-as-decreasing-returns-to-scale.

Koopmans, Tjalling C. "Three Essays on the State of Economic Science." (1959).

Magill, Michael, and Martine Quinzii. *Theory of Incomplete Markets*. Vol. 1. Mit press, 1996.

Vazirani, Vijay V. "Nonseparable, Concave Utilities Are Easy---in a Perfect Price Discrimination Market Model." *SIAM Journal on Discrete Mathematics* 27.1 (2013): 266-273.

Appendix for Chapter 3

Proof of Second Welfare Theorem:

I now fill in the gaps of the proof left from the outline presented in the main body of the paper. This outline had seven steps.

1. V is an open and convex set.

To see that V is open, consider any c'_θ $\forall\theta$ such that $c'_\theta(i) \succcurlyeq_\theta c_\theta(i)$ $\forall i$ $\forall\theta$, and furthermore for some type θ^*:

$$\mathbb{P}\Big(c'_{\theta^*}(i) \succ_{\theta^*} c_{\theta^*}(i)\Big) > 0$$

Pick i^* such that c'_{θ^*} and c_{θ^*} are continuous at i^*, and $c'_{\theta^*}(i^*) \succ_{\theta^*} c_{\theta^*}(i^*)$. Note that continuity of preferences implies that there is some value $\delta > 0$ such that in an open neighborhood N^* around i^*:

$$\|\tilde{c} - c'_{\theta^*}\| < \delta \Rightarrow \tilde{c} \succcurlyeq_{\theta^*} c_{\theta^*}(i)$$

Define $\epsilon \equiv \delta\mathbb{P}(N^*)$, and note that $\epsilon > 0$. Pick arbitrary $\tilde{C}$ such that:

$$\left\|\tilde{C} - \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di\right\| < \epsilon$$

Define $\tilde{c}_{\theta^*}(i) \equiv c'_{\theta^*}(i) + [\mathbb{P}(N^*)]^{-1}\left[\tilde{C} - \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di\right]$ for those agents $i \in N^*$.

Since $\left\|[\mathbb{P}(N^*)]^{-1}\left[\tilde{C} - \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di\right]\right\| < \delta$, we know that $\tilde{c}_{\theta^*}(i) \succ_{\theta^*} c_{\theta^*}(i)$ for these same agents $i \in N^*$. Defining $\tilde{c}_\theta(i) = c_\theta(i)$ otherwise, we get that:

$$\tilde{C} = \sum_{\theta} \lambda_{\theta} \int_{0}^{1} \tilde{c}_{\theta}(i)di \in V$$

Thus, we know any element of V has an ϵ-ball around that element also in V.

That V is convex is a corollary of Liapunov's Theorem, as demonstrated in Hildenbrand (1974, pg. 62 theorem 3 and the following corollary).

2. The same result from Hildenbrand (1974) also demonstrates that R is convex.

   Alternatively, one need only note that under assumption (1):

$$R = \sum_{\theta} \lambda_{\theta} \big[\{\omega_{\theta}\} + conv\big(\tilde{Y}_{\theta}\big)\big]$$

This uses the fact that elements of $\mathcal{J}$ cannot be produced, so that the constraint $c_{\theta l} = \omega_{\theta l} + y_{\theta l}\ \forall l \in \mathcal{J}$ has no implications for aggregate consumption.

3. This step can be decomposed into two parts:

   a. If $V = \emptyset$, then equilibrium arises from $p = \mathbf{0}$.

First, I want to show that $V = \emptyset$ implies that a.e. agent is globally sated. To show the contrapositive, suppose that $\exists\tilde{\theta}$ such that $\exists c'_{\tilde{\theta}}(i) \succ_{\tilde{\theta}} c_{\tilde{\theta}}(i)$ for uncountably many $i \in [0,1]$. Then $c_{\tilde{\theta}}$ is continuous at some such value $\tilde{\imath}$, so that continuity of preferences implies that for some open neighborhood $\tilde{N}$ around $\tilde{\imath}$:

$$c'_{\tilde{\theta}}(\tilde{\imath}) \succ_{\tilde{\theta}} c_{\tilde{\theta}}(i)\ \forall i \in \tilde{N}$$

This implies that V is not empty:

$$\sum_{\theta \neq \tilde{\theta}} \lambda_{\theta} \int_{0}^{1} c_{\theta}(i)di + \lambda_{\tilde{\theta}} \left[\int_{\tilde{N}} c'_{\tilde{\theta}}(\tilde{\imath})di + \int_{[0,1]\setminus\tilde{N}} c_{\tilde{\theta}}(i)di\right] \in V$$

By the contrapositive, $V = \emptyset$ implies that $((c_\theta, y_\theta)_\theta, \mathbf{0})$ is an equilibrium. We assume for the rest of this proof that $V \neq \emptyset$.

b. If $V \neq \emptyset$, then $\exists p \in \mathbb{R}^L,\ v \in \mathbb{R}$ such that $(X, C) \in R \times V \Rightarrow p * X \leq v < p * C$.

In the body of the paper, I noted that by the Separating Hyperplane Theorem.

$$\exists p, v \colon (X, C) \in R \times V \Rightarrow p * X \leq v \leq p * C$$

Furthermore, $p \neq \mathbf{0}$. Suppose for the sake of contradiction that $\exists C \in V \colon p * C = v$. Then for any element $l^* \in \{1, \dots, L\}$ such that $p_{l^*} \neq 0$, one can consider $\tilde{C} \equiv (C_l - sign(l^*)\mathbb{I}(l = l^*)\epsilon)_l$ for sufficiently small ϵ so that $\tilde{C} \in V$, yet $p * C < v$. This contradicts the claim of the Separating Hyperplane Theorem. Therefore:

$$\exists p \neq \mathbf{0}, v \in \mathbb{R} \colon (X, C) \in R \times V \Rightarrow p * X \leq v < p * C$$

4. $\sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \in cl(V)$

I reuse notation from the argument from step 1, since the reasoning is similar. Since V is not empty, there is an uncountable number of points i at which $c_{\theta^*}(i) \notin \underset{\mathbb{R}_+^L}{\operatorname{argmax}} \succcurlyeq_{\theta^*}$ for some $\theta^* \in \Theta$. For some such i^*, c_{θ^*} is continuous at i^*. Pick $c^* \succ_{\theta^*} c_{\theta^*}(i^*)$. Then there is an open neighborhood N^* such that $c^* \succ_{\theta^*} c_{\theta^*}(i)\ \forall i \in N^*$. For any $\epsilon > 0$, define $N_\epsilon^* \subset N^*$ such that when defining:

$$c_\theta^\epsilon(i) \equiv \mathbb{I}(\theta = \theta^*, i \in N_\epsilon^*)c^* + \mathbb{I}(\theta \neq \theta^* \text{ } or \text{ } i \notin N_\epsilon^*)c_\theta(i)$$

We obtain:

$$\left\| \sum_\theta \lambda_\theta \int_0^1 c_\theta^\epsilon(i)di - \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \right\| < \epsilon$$

Yet $\sum_\theta \lambda_\theta \int_0^1 c_\theta^\epsilon(i)di \in V$. Thus, $\sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \in cl(V)$.

5. The aggregated allocation $\sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di$ is on the hyperplane.

Note that $p * \sum_\theta \lambda_\theta \int_0^1 c_\theta^\epsilon(i)di \geq v\ \forall \epsilon > 0$. By continuity of the dot product:

$$p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \leq v \leq \lim_{\epsilon \to 0^+} p * \sum_\theta \lambda_\theta \int_0^1 c_\theta^\epsilon(i)di = p * \lim_{\epsilon \to 0^+} \sum_\theta \lambda_\theta \int_0^1 c_\theta^\epsilon(i)di$$

$$= p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di$$

Thus, $p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \leq v \leq p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di \Rightarrow p * \sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di = v$.

6. $T_\theta(i) \equiv p * [c_\theta(i) - s_\theta(i)] \Rightarrow \sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = 0$

First, note that $T_\theta(i)$ is integrable. Second, note that:

$$\sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di = p * \left[\sum_\theta \lambda_\theta \int_0^1 c_\theta(i)di - \sum_\theta \lambda_\theta \int_0^1 [\omega_\theta + y_\theta(i)]di\right] = p * \mathbf{0} = 0$$

7. $((c_\theta, y_\theta)_\theta, p)$ is an equilibrium with integrable transfers T_θ: $[0,1] \to \mathbb{R}$.

Suppose that $((c_\theta, y_\theta)_\theta, p)$ violates the preference maximization requirement of equilibrium.

Then $\exists \theta^*$ such that for uncountably many agents i:

$$c_{\theta^*}(i) \notin \underset{c \in B_{\theta^*}\left(p, \omega_{\theta^*}, T_{\theta^*}(i^*)\right)}{\operatorname{argmax}} \succcurlyeq_{\theta^*}$$

There is some such i^* at which c_{θ^*} and T_{θ^*} are continuous.[18] Then for some neighborhood N^* around i^*:

$$c_{\theta^*}(i) \notin \underset{c \in B_{\theta^*}\left(p, \omega_{\theta^*}, T_{\theta^*}(i)\right)}{\operatorname{argmax}} \succcurlyeq_{\theta^*}\ \ \forall i \in N^*$$

Since N^* has positive measure, it follows that the set $\left\{c_{\theta^*}(i) \notin \underset{B_{\theta^*}\left(p, \omega_{\theta^*}, T_{\theta^*}(i)\right)}{argmax} \succcurlyeq_{\theta^*}\right\}$ contains a set of positive measure.

[18] This uses the fact that T_θ is continuous a.e. $\forall \theta$.

One can pick $\left(c'_{\theta^*}(i), y'_\theta(i)\right)$ such that for all $i \in N^*$:

$$c'_{\theta^*}(i) \succ_{\theta^*} c_{\theta^*}(i)$$

$$c'_{\theta^*}(i) \in B_\theta\left(p, \omega_{\theta^*}, y'_{\theta^*}, T_{\theta^*}(i)\right)$$

$$y'_{\theta^*}(i) \in Y_\theta$$

Otherwise, let $\left(c'_\theta(i), y'_\theta(i)\right) = \left(c_\theta(i), y_\theta(i)\right)$. As described above, this yields contradiction:

$$v < p * \sum_\theta \lambda_\theta \int_0^1 c'_\theta(i)di \leq p * \sum_\theta \lambda_\theta \int_0^1 s'_\theta(i)di + \sum_\theta \lambda_\theta \int_0^1 T_\theta(i)di \leq v$$

Thus, we know that all consumers maximize preferences subject to their budget constraints. ■