The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards

Koichiro Ito
Boston University and NBER

James M. Sallee*
University of Chicago and NBER

December 24, 2013

Abstract

In many countries, fuel-economy standards mandate that vehicles meet a certain fuel economy, but heavier or larger vehicles are allowed to meet a lower standard. This has the perverse implication of allowing automakers to meet standards either by improving fuel economy or by increasing weight, which lowers fuel economy and increases externalities related to accidents. This is but one example of an attribute-based regulation, in which the subsidy, tax or regulation imposed on a product is a function not just of the amount of an externality that the product generates, but also how each product's externality compares to that of other products deemed to be similar by virtue of a commonality in some other attribute. Such policies are ubiquitous, but the core logic and welfare consequences of their deployment have not been studied by academic economists. This paper develops an analytical framework that captures the central implications of attribute-based policies, characterizes the deadweight loss caused by attribute-basing, and establishes situations in which attribute-basing may be efficient. The paper then empirically examines the consequences of attribute-based fuel economy standards in Japan, where fuel economy standards are an attribute-based function of vehicle weight. We use cross-sectional and panel techniques to demonstrate that attribute-based regulation has significantly altered the distribution of vehicle weight in Japan. For cars whose weight is altered in response to the policy, we estimate that the alteration generates a welfare loss from the exacerbation of weight-related externalities of $1441 per unit sold, which translates into a $648 million annual loss across the Japanese auto market.

Keywords: fuel economy standards, corrective taxation, notches
JEL: H23, Q48, L62

*Ito: ito@bu.edu. Sallee: sallee@uchicago.edu. The authors would like to thank Kunihito Sasaki for excellent research assistance. For helpful comments, we thank Soren Anderson, Damon Jones, Richard Newell and seminar participants at Michigan State University and the University of Chicago. ©2013 Koichiro Ito and James M. Sallee. All rights reserved.

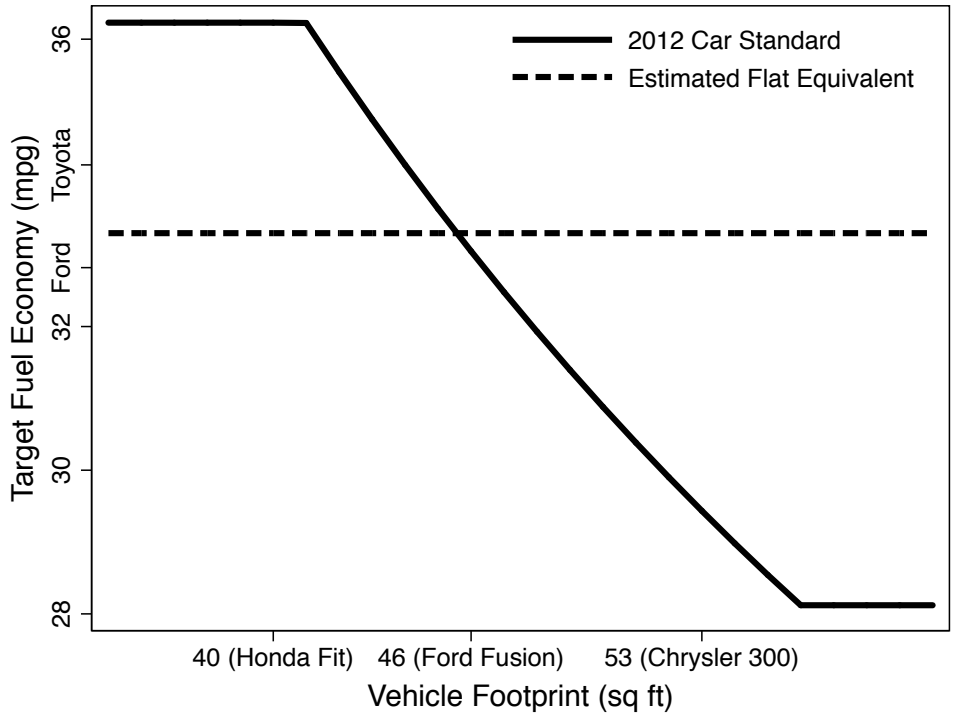
1 Introduction

The goal of this paper is to advance economists' understanding of a feature called attribute-basing that is common to many policies aimed at correcting externalities but is apparently at odds with the basic logic of Pigouvian taxation. Attribute-based policies are policies in which the corrective policy—which could be a regulation, a tax or a subsidy—imposed on a product is not simply a function of how much externality the product creates, but instead depends on how much externality the product creates relative to some benchmark, which itself depends on some other attribute of the product. Attribute-basing is common to corrective policies, in particular policies aimed at improving the energy efficiency of energy consuming durable goods, but there is no academic literature that explores the optimality, merits and problems with such policies.

Attribute-basing is perhaps best explained by example. Corporate Average Fuel Economy (CAFE) standards in the United States became attribute-based as of 2012. Prior to 2012, every automaker had to sell vehicles that met or exceeded a fleet-average fuel economy of some particular level, which was common to all producers. Since 2012, CAFE specifies a schedule of fuel economy targets for each vehicle that is a declining function of the vehicle's "footprint"—the rectangular area inside of the vehicle's tires. Each automaker has to meet a standard equal to the average target of the vehicles that it sells, so that automakers who sell larger vehicles on average are permitted to have less efficient vehicles. Automakers who exceed their standard are permitted to sell their excess credits to other automakers, so that the policy ultimately imposes a single minimum on the entire market.

Figure 1 illustrates this schedule for 2012 and shows the large difference in targets across vehicles and automakers. The Honda Fit (a compact car) has a target fuel economy of 36.2 miles per gallon, whereas the Chrysler 300 (a full-size sedan) has a target of 29.4. The steepness of this schedule creates notable differences in the fuel economy required of different automakers who produce cars of different average size. Ford and Toyota, whose targets differ by 1.4 miles-per-gallon in 2012, are labeled in the figure. Vehicle footprint is not immutable, however, so the footprint-based standard not only creates dispersion in requirements across automakers, but also creates an incentive for automakers to enlarge footprint, which runs counter to the goals of the policy by endorsing larger (and hence less energy efficient) vehicles.

Figure 1: Example of an attribute-based standard: 2012 U.S. new car CAFE



As of 2011, the CAFE target fuel economy for a vehicle depends on its footprint. Three footprint values are labeled with example vehicles that have that footprint. The flat equivalent is the EPA's estimate based on the distribution of footprint in current fleet. Prior to 2011, the new car CAFE standard was a flat standard at 27.5 mpg. The firm specific target values for Ford and Toyota based on model year 2011 average footprint is labeled. Data taken from the Federal Register, 40 CFR Part 85.

Such attribute-basing is ubiquitous in the realm of energy efficiency policies. Appliance standards are typically a function of the size or features of the appliance. The same is true for the cutoff value to qualify an appliance for Energy Star certification, which is frequently also used as the criteria for appliance rebate programs meant to conserve energy. The hybrid vehicle tax credit was attribute-based because the subsidy amount was based on each hybrid's gasoline conservation relative to a benchmark vehicle, determined by its market segment. The U.S. only recently moved to attribute-basing for CAFE, but such policies have long-existed in other countries. For example, the Chinese fuel economy standard is an attributed-based function of engine displacement. In Japan, the regulation and government subsidy programs are sliding scales of a vehicle's weight. Circulation and registration taxes in Germany, France, Sweden and the United Kingdom depend

on a vehicle's engine size. Label ratings, which identify a vehicle on a scale of most to least polluting in various places, including Europe and Korea, depend on vehicle size. All of these policies provide a *lower* standard for products that are larger and *more polluting*, which generates perverse incentives.[1]

In spite of the frequent appearance of attribute-basing in actual policies, the academic literature has provided little scrutiny of its potential costs and benefits. To the extent that these policies are aimed at correcting externalities related to energy consumption, attribute-basing would seem to have no role in an optimal policy because the first-best Pigouvian tax on energy consumption does not depend on anything other than marginal damages. Thus, on its face, attribute-based regulation appears to be superfluous in terms of achieving efficient changes in energy consumption, and it creates an undesirable byproduct via the distortion of the choice of the attribute itself.

Might there be additional considerations that make attribute-basing a good idea? If so, what would characterize the optimal design of an attribute-based policy? If not, so that attribute-basing is inefficient, what distortions does it create and what determines the size of any associated deadweight loss? Is there empirical evidence that distortions exist?

This paper aims to make an important first step in the analysis of attribute-basing by proposing answers to these questions and generally evolving our understanding of this class of public policy. We begin in section 2 by establishing an analytical framework that considers two distinct cases, which differ in whether or not the attribute-based regulation allows compliance trading. For policies that allow compliance trading—that is, firms that exceed their standard can sell their excess compliance to other firms who fall short—a regulation imposes a uniform shadow price on all market actors. An example of such a policy is CAFE. In this case, attribute-based regulation is identical to a uniform corrective tax, and the logic of Pigou prevails: attribute-basing is unnecessary and distortionary.

The use of attribute-based regulation in this case creates several distortions. First, it creates a direct distortion analogous to a Harberger triangle in the choice of the attribute. It may also induce an indirect distortion in the choice of energy efficiency. These distortions are "second-order" in the size of the distortionary tax wedge in that excess burden is a function of the tax squared. The

[1] Counterexamples include circulation and registration taxes for vehicles in Europe. In several countries, these taxes are an increasing function of vehicle characteristics, like cylinders or weight, that are positively correlated with pollution.

existence of an attribute-based target causes the second-best tax rate on energy consumption to be below the first-best Pigouvian rate, with the difference between the first and second-best taxes depending on the proportion of compliance that comes from raising the attribute versus improving energy efficiency. Attribute-basing will also create "first-order" (linear in the tax) distortions when the attribute itself is associated with an unpriced externality. This applies to the case of automobiles when the attribute is weight or footprint because those characteristics are correlated with safety externalities (Anderson and Auffhammer Forthcoming).

In the second case, when compliance trading is not allowed, all of the same distortions exist, but there are potential benefits from marginal cost equalization. If trading is not allowed, then a uniform regulation that forces all products to meet the same energy efficiency minima will generate difference costs of compliance. This is equivalent to imposing different corrective taxes on different products, which leads to an inefficient dispersion in the equilibrium marginal cost of compliance. Attribute-basing can reduce the dispersion in marginal costs to the extent that marginal costs are correlated with the attribute. We show that the the slopes of actual attribute-based policies are made to match existing data, which is the second-best optimal solution in the special case where all products have a common cost of improving energy efficiency and attributes are exogenous. In this case, the R^2 from a regression of energy consumption on the attribute is a summary statistic that captures the relative efficiency of attribute-basing as compared to trading.

The paper then focuses on empirical evidence from Japanese fuel economy standards, which are particularly amenable for analysis, in section 3. Japanese fuel economy standards require that automakers meet a minimum fuel economy standard, but the standard depends on each vehicle's weight, with heavier vehicles needing to meet laxer standards. Moreover, the minima are discrete functions of weight class, so that vehicles on either side of a weight threshold are required to meet discretely different standards. These "notches" provide a straightforward way to identify distortions in the attribute in response to policy.

Cross-sectional data provide clear evidence of bunching in the weight distribution. We estimate that around 10% of Japanese vehicles have an increase in weight by an average of 104 kilograms as a result of the weight-basing used in the fuel economy regulations. We corroborate these estimates with panel data, and we also estimate an elasticity of the change in weight with respect to the fuel economy standard's stringency and conclude that a one percent increase in the stringency of

the standard leads to a .16 percent increase in vehicle weight. Relating these estimates to our theoretical results, we conclude that the second-best subsidy on fuel economy in Japan is 40% lower than the first-best, as a result of distortions in weight. We also estimate a lower bound on the welfare distortions from the existing policy by quantifying the first-order distortion due to exacerbation of unpriced safety externalities. Based on recent estimates of the value of externalities related to vehicle weight, we estimate that this creates welfare losses on the order of $1441 per car sold, which aggregates to roughly $648 million annually across the new car market in Japan.

To the best of our knowledge, we are the first paper to take an analytical approach to assessing the merits and distortions caused by attribute-basing in policies aimed at correcting externalities. Our work is nevertheless related to several strands of literature. The literature on the move to footprint-based CAFE rules in the U.S. includes Whitefoot and Skerlos (2012), which uses engineering estimates of design costs and a discrete-choice economic model to predict how much automakers will manipulate footprint in response to a tightening of CAFE standards. That study concludes that an increase in footprint will be a major source of adjustment, which is broadly consistent with our findings of considerable manipulation of weight in the Japanese car market. Gillingham (2013) discusses the implicit incentive for the expansion of footprint in a broader discussion of CAFE policies. Jacbosen (2013) addresses the safety impacts of footprint-based standards in the U.S., which our work will also touch on. Our consideration of notched attributed-based policies relates to the literature on notched corrective taxation, which began with Blinder and Rosen (1985), and includes prior analysis of automobile fuel economy in Sallee and Slemrod (2012).

2 Theoretical framework

The theoretical portion of this paper explores the use of attribute-based regulations for correcting externalities. We first present our notation and define our setup. In section 2.1 we back up and discuss the economic content of our key modeling assumptions before proceeding to the model's solutions.

We model a consumer who buys a single unit of a durable good that has two characteristics, one that is related to an externality (e.g., energy efficiency), denoted e, and some other attribute (e.g., footprint), denoted a. The consumer pays price $P(a, e)$ for one unit of the durable and spends the

remainder of his/her exogenous income I on an outside good x, the price of which we normalize to 1 so that it is the numeraire.

Consumer utility is determined by a flow utility from the durable $U(a, e)$, the utility from consumption of all other goods, denoted $h(x)$, and an externality $\phi(E)$, where E is the sum of e over all consumers. We are not particularly interested in complementarities of consumption between the energy-consuming durable and the remainder of the consumer's consumption bundle (or the externality), so we will assume that utility is separable in the three components. Consumer welfare is thus $W = U(a, e) + h(x) + \phi(E)$.[2]

Given our assumption of quasilinearity, the marginal utility of income will be the consumption value of the numeraire h'. To simplify notation, we assume that there is a constant marginal utility of income and we scale utility by this value. Likewise, for notational convenience, we assume that the externality has a constant marginal damage, so that $\phi(E) = \phi \cdot E$. We then rewrite consumer welfare as $W = U(a, e) + x + \phi \cdot E$.

The government can levy a tax on the durable that is a function of both attributes, denoted $t(a, e)$. Our model will focus on the case with a linear tax on e and an attribute-based target function, which implies that $t(\cdot)$ will have the form $t(e - \sigma(a))$. All revenue is recycled through a lump-sum demogrant G. In all cases we assume that there are many individual consumers and each therefore ignores their own role in creating the externality—i.e., they take E as fixed. The consumer's optimization problem is then:

$$\begin{aligned} \max_{a,e,x} W =& U(a, e) + x + \phi \cdot E \\ & \text{s.t. } I + G \geq P(a, e) + t(a, e) + x. \end{aligned}$$

Note that this differs from a standard consumer choice problem over the three goods a, e and x only in that the price and tax paid for a and e are possibly related to each other and therefore nonlinear. This is the complete representation of the demand side of the market.

On the supply side, we assume perfect competition. Specifically, we assume that there is a constant marginal cost of producing a product with attribute bundle (a, e), which we denote as

[2]Note that it is reasonable to model energy efficiency as effecting utility only through cost savings that enable greater consumption of the numeraire. We use the more general form here in the hopes that our model be useful beyond the case when e denotes energy efficiency.

$C(a, e)$, and that there are no fixed costs or barriers to entry. This means that (in the absence of policy) a consumer will be able to purchase any bundle of attributes (a, e) and pay $P(a, e) = C(a, e)$. Furthermore, because there are no profits in equilibrium due to free entry and the lack of fixed costs, firms will make no profits and consumer surplus will be a sufficient welfare statistic. We assume that $C(a, e)$ is rising and convex in both attributes. In all real world policies that we know of, the attribute a—when written as a good as opposed to a bad—is negatively related to energy efficiency e. (For cars, a is either footprint or weight, both of which correlate negatively with fuel economy.) Thus, we assume that $\partial^2 C/\partial a \partial e \geq 0$. This ensures the convexity of the budget set.[3]

2.1 Discussion of the model's key assumptions

Before discussing the model's solutions, we briefly discuss several key modeling assumptions that we employ, namely the relationship between e and the externality, perfect competition, and the use of a tax policy instead of a regulation.

First, our model assumes that the externality is directly (and linearly) related to the energy efficiency, e. In reality, externalities from energy consumption depend on energy efficiency and utilization, which is itself a function of efficiency. (For example, externalities related to climate change from gasoline consumption depend on vehicle fuel economy and miles driven, and miles driven depends on fuel economy.) This induces what is commonly referred to as a rebound effect.[4] Moreover, policies that target only the new product market induce offsetting effects in the used durable market.[5] For these reasons, it is well understood that energy efficiency policies—whether they are regulatory policies like CAFE or equivalent taxes on the products—represent significant departures from a first-best Pigouvian tax on the externality itself.[6] Our focus is on the marginal efficiency impacts of attribute-basing. For that reason, we will abstract from these considerable concerns and assume that there is some externality related directly to energy efficiency e and that a tax on that feature of a product represents a first-best solution.

Second, our assumption of perfectly competitive supply abstracts from potentially interesting

[3]Empirical estimates in Knittel (2011) indicate that the log of fuel economy is concave down in the log of weight, which is the specific application we have in mind.

[4]The margin of adjustment described here is sometimes called the "direct rebound" effect. There are also other effects that work through income. See Borenstein (2013) for a recent discussion.

[5]See Jacobsen and van Benthem (2013) for an exploration of this effect for the case of automobiles.

[6]See Anderson, Parry, Sallee, and Fischer (2011) for a recent review of CAFE that describes related evidence. For the equivalence of vehicle taxes and CAFE, see Sallee (2011) and Gillingham (2013).

questions regarding how the presence of attribute-based incentives influences welfare through its impact on mark-ups, product differentiation and competition among firms. Despite these limitations, we assume perfect competition to facilitate the comparison between attribute-basing and Pigouvian taxation and to make the welfare analysis transparent and tractable. We believe that our approach will provide a useful foundation upon which to build a subsequent study of attribute-basing that focuses on imperfect competition. Moreover, some studies of CAFE take a short-run perspective, in which the set of vehicles sold is fixed and firms must improve fuel economy by modifying the sales shares of their models. In this sense, perfect competition naturally provides an analysis of the long-run, in which no product characteristics are fixed.

Third, many of our examples regard regulations that, like CAFE, apply to an average across multiple products sold by the same firm. The standard way of modeling firms with multiple products under imperfect competition implies that firms must worry about how increasing some attribute of one of their products will affect the demand for all of their other products. Our assumption about perfect competition abstracts from those concerns because, when firms make zero profits and consumers can purchase any bundle of attributes, there will be no lost profits from cannibalization. Thus, under perfect competition, nothing is lost by assuming that firms produce a single product, which is the viewpoint we adopt.

Fourth, we model a tax policy instead of a regulation. In appendix A we demonstrate that we can write a tax policy that is equivalent to a regulation that requires e to exceed a minimum value of σ. Call the shadow price of the regulation ν. A tax on e that equals $t(e-k)$ is equivalent to the regulation when $t=\nu$ and $k=\sigma$. Thus, to translate our results, we interpret our tax rate as the shadow price in the equivalent regulation. In appendix A we show that the equivalence extends to attribute-based policies. If the attribute-based regulation requires a product's e to exceed $\sigma(a)$ and this yields a shadow price of ν, then a tax of $t(\omega(a)-k)$ is equivalent to the regulation when $\sigma(a)=\omega(a)$ and $t=\nu$.[7] (These results are true for multi-product or single-product firms.) Given this equivalence, we focus primarily on tax policies, which link more directly to the public finance literature's results on distortionary wedges and Pigouvian taxation.

[7]When $k \neq 0$, regulatory policies differ from the Pigouvian tax in that they provide different incentives on the extensive margin. That margin is assumed away in this context, though it has been shown to be relevant in other settings, for example in Holland, Knittel, and Hughes (2009). This margin is why taxes on a product's energy consumption are not equivalent to subsidies for its energy efficiency (Metcalf 2009).

Fifth and final is our assumption regarding trading. Some regulations, including CAFE, allow firms to trade excess compliance credits across products or firms. If Toyota's fuel economy exceeds its target, for example, it may sell its excess compliance to BMW, who can credit their account accordingly. When full trading exists, the shadow prices of the regulation will be equalized across all producers. In that case, the equivalent tax model imposes a single tax rate on all firms. In some other policies, there is no trading. Appliance standards, for example, generally require each product to meet the regulatory minimum individually. In this case, the individual products may face different shadow prices, and the equivalent tax policy would be one in which tax rates differ across products. We will consider each of these cases separately, starting with the case of trading.

2.2 Baseline model with trading

We begin by considering the case with trading, so that there is a single shadow price (tax rate) facing all consumers. In this case, we focus on the problem with only one type of consumer, which provides great notational simplification. (If there were many types of consumers, their problems would be separable because of our perfect competition assumptions and our solutions would differ only in featuring summations over types.) We further assume that there are measure 1 consumers, so that $E = e$.

2.2.1 First-best solutions do not involve attribute-basing

The first step in our analysis is to derive the first-order conditions for the consumer facing a tax policy and compare them to the planner's first-best conditions when the planner is allowed to choose the allocation directly. We can then see what function $t(a, e)$ would make the consumer's conditions equal to the first-best.

Under these assumptions, we now write the consumer's problem as an unconstrained optimization problem by substituting the budget constraint into the objective function:[8]

$$\max_{a,e} W = U(a, e) + [I + G - P(a, e) - t \cdot (e - \sigma(a))] + \phi \cdot E.$$

[8] Our assumptions regarding $C(a, e)$ ensure that the budget set is convex and thus that the consumer will exhaust their budget.

This is a standard choice problem, and the first-order conditions can be derived directly by differentiation:

$$\frac{\partial W}{\partial a} = \frac{\partial U}{\partial a} - \frac{\partial P}{\partial a} + t\sigma' = 0 \tag{1}$$

$$\frac{\partial W}{\partial e} = \frac{\partial U}{\partial e} - \frac{\partial P}{\partial e} - t = 0, \tag{2}$$

where σ' is the derivative of $\sigma(a)$ with respect to a.

First-order conditions for the first-best allocation can be found by differentiating the planner's problem, which is:

$$\max_{a,e} SWF = U(a,e) + [I - C(a,e)] + \phi \cdot e.$$

This differs from the consumer's problem in directly inserting the cost function in place of prices (though these will be equal under perfect competition) and taxes, and in recognizing how the choice of e affects the total externality. The first-best optimization conditions are found by differentiation:

$$\frac{\partial SWF}{\partial a} = \frac{\partial U}{\partial a} - \frac{\partial C}{\partial a} = 0 \tag{3}$$

$$\frac{\partial SWF}{\partial e} = \frac{\partial U}{\partial e} - \frac{\partial C}{\partial e} + \phi = 0. \tag{4}$$

Under perfect competition, prices reflect unit costs, so $\frac{\partial P}{\partial a} = \frac{\partial C}{\partial a}$ and $\frac{\partial P}{\partial e} = \frac{\partial C}{\partial e}$. Then, it is immediately apparent that the only way to make the consumer's first-order conditions equal to the first-best (that is, to make equation 1=equation 3 and equation 2=equation 4) is to set $\sigma' = 0$ and $t = -\phi$. That is, to have a Pigouvian subsidy on e with no attribute-basing.

This is the first result of our theoretical analysis. The result is not surprising, but it is relevant for evaluating policy, as it establishes a set of conditions under which attribute-basing is counterproductive. If the only market failure is an externality associated with e, then the optimal policy will not involve attribute-basing. The optimal attribute-based slope is uniquely zero. This result will be unchanged if there are multiple types of consumers, and it translates directly into the result that there should be no attribute-basing in a regulation, provided that there is trading. If there is a role for attribute-basing, it must be due to some additional factor. We discuss several possibilities

below.

2.2.2 Attribute-basing causes deadweight loss

Suppose that the true welfare model is as specified above and that the tax rate is set at the Pigouvian level, $t = -\phi$. The optimal attribute slope is zero, but suppose the policy is attribute-based, $\sigma' \neq 0$, for some outside reason, such as political economy or simple confusion on the part of policymakers. What distortions does attribute-basing cause in this setting?

A qualitative characterization of distortions can be made by examining the first-order conditions of the consumer, again comparing this to the first-best. Let a^* and e^* denote the first-best choices. Let a' denote the value of a that is chosen under the attribute standard. Then the following are the first-order conditions for e from the two problems:

$$\begin{aligned} \text{First-best:} \quad & \frac{\partial U(a^*, e^*)}{\partial e} + \phi = \frac{\partial P(a^*, e^*)}{\partial e} \\ \text{Consumer's choice:} \quad & \frac{\partial U(a', e')}{\partial e} - t = \frac{\partial P(a', e')}{\partial e}. \end{aligned}$$

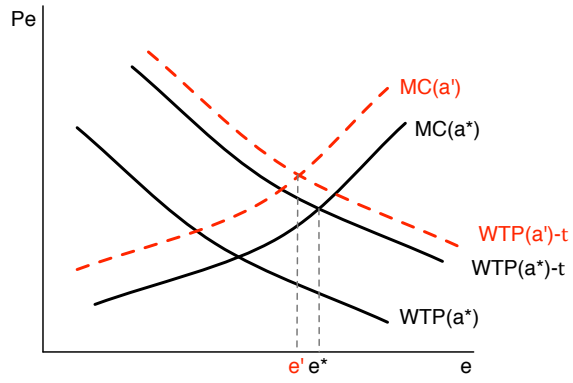
If $t = -\phi$, as would be the case if policy-makers used the Pigouvian tax level in the attribute-based problem, then the consumer's first-order condition is exactly the same as the first-best, except that the level of a' is distorted. Thus, there is no "partial equilibrium" distortion in the choice of e, but there may be a "general equilibrium" distortion in e.

This is illustrated in Figure 2a, which shows the consumer's choice of e by plotting the marginal cost and marginal benefit of (willingness to pay for) e, conditional on $a = a^*$. (We do not need to condition on x because of the quasi-linearity of $U(a, e)$.) The marginal cost of increasing e to the consumer is the derivative of the cost function with respect to e, given a level of a. The private marginal benefit of increasing e is the derivative of U with respect to e, for a given level of a. The solid black lines plot hypothetical curves for these functions when $a = a^*$. The corrective tax rate t shifts the willingness to pay curve up by amount t. Thus, the consumer will choose e^* when there is no attribute-basing and the tax rate is t.

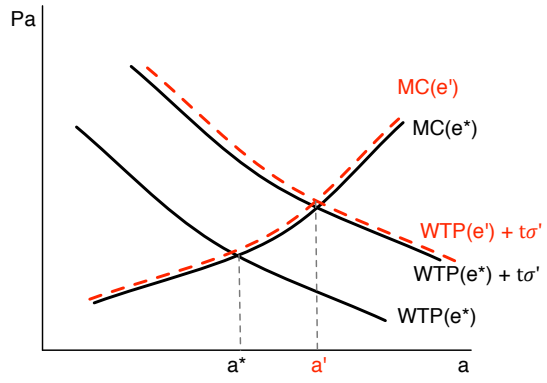
Suppose that, under attribute-basing, the consumer's choice of a, call it a', exceeds a^*. In the diagram, this will shift both the marginal cost and the marginal benefit curves, depending on the cross-partial derivatives of price and flow utility. The marginal cost curve will shift up by

Figure 2: Distortions in the choice of a and e from attribute-basing

(a) Choice of e (second-order distortion)



(b) Choice of a (first-order distortion)



$\approx (a' - a^*)\frac{\partial^2 P(a^*,e^*)}{\partial a \partial e}$. The willingness to pay curve will shift by $\approx (a' - a^*)\frac{\partial^2 U(a^*,e^*)}{\partial a \partial e}$. If $\frac{\partial^2 P(a^*,e^*)}{\partial a \partial e} = \frac{\partial^2 U(a^*,e^*)}{\partial a \partial e}$ (which would occur only by coincidence) then $e^* = e'$ and attribute-basing would cause no change in the allocation of e. Physics suggests that the cross-partial is positive for the cost function, but the sign of the cross-partial for utility is ambiguous. If it is positive, then the two effects will be offsetting. If it is negative or zero, then the general equilibrium effect on e will unambiguously lead to a decrease in e.

The same is not true in the analysis of a. The first-order conditions with respect to a are:

$$\begin{aligned} \text{First-best:} \quad & \frac{\partial U(a^*, e^*)}{\partial a} = \frac{\partial P(a^*, e^*)}{\partial a} \\ \text{Consumer's choice:} \quad & \frac{\partial U(a', e')}{\partial a} + \underbrace{t\sigma'}_{>0} = \frac{\partial P(a', e')}{\partial a}. \end{aligned}$$

When $\sigma' \neq 0$, attribute-basing introduces a wedge between the marginal benefit and marginal cost functions for a. This is directly analogous to a Harberger triangle in standard partial equilibrium tax analysis, where the tax wedge is of size $\sigma' t$. The deadweight loss due to the price wedge will be $\approx 1/2 \cdot (a' - a^*)\sigma' t$. This approximated loss is "second-order" in the same sense that the canonical Harberger triangle is; it is a function of $(\sigma' t)^2$ (because $a' - a^*$ is locally linear in $\sigma' t$).

The difference between e^* and e' will also cause a change in the choice of a. This will also be a second-order effect, of magnitude $\approx (e' - e^*)\left(\frac{\partial^2 P(a^*,e^*)}{\partial a \partial e} - \frac{\partial^2 U(a^*,e^*)}{\partial a \partial e}\right)$. The direct effect of attribute-based standards is in creating the wedge $-t\sigma'$, which equals $\phi\sigma'$ when t is set at the

Pigouvian level. This wedge will cause the consumer to choose too high of a value of a, and the distortion will be rising as σ gets steeper (σ' gets more negative). The wedge is also larger (in absolute value) as t gets larger (or as ϕ gets larger assuming the Pigouvian rate for t is used).

This is shown in Figure 2b, which plots the marginal cost and benefit of a, conditional on a given level of e. The introduction of the attribute-based wedge shifts the benefit curve out linearly with σ'. This creates what would be labeled as a Harberger triangle in a standard partial equilibrium graph, showing the lost utility from choosing the too large value of a'.

This analysis has important implications for empirical studies of attribute-based policies. Take CAFE for example. The introduction of attribute-basing may have only a very limited impact on fuel economy as compared to a policy with the same implicit subsidy for fuel economy that has no attribute-basing. Fuel economy might be higher or lower under the attribute-based policy than under the shadow-price equivalent attribute-free policy. The first-order distortions and welfare losses of the footprint-based standards are associated with the distortion in the footprint, not the distortion in fuel economy. The size of the externality enters the distortion formulas because it determines the size of the tax wedge that distorts the attribute, not because the primary welfare loss is from too little externality reduction. Altogether, this suggests that the place to look for distortions in the data is in the distribution of the attribute upon which standards are based, not in the distribution of energy efficiency.

2.2.3 Attribute basing reduces the second-best tax on the externality

The analysis above assumed that the tax rate was set at the Pigouvian rate and σ' was taken as given. An alternative is to ask what the second-best t will be, taking σ' as given. The planner's problem is:

$$\max_t U(a,e) - P(a,e) + \phi e.$$

The planner's first-order condition is:

$$\left(\frac{\partial U}{\partial e} - \frac{\partial P}{\partial e} + \phi\right)\frac{\partial e}{\partial t} + \left(\frac{\partial U}{\partial a} - \frac{\partial P}{\partial a}\right)\frac{\partial a}{\partial t} = 0.$$

Substituting in the first-order conditions of the consumers yields:

$$(t+\phi)\frac{\partial e}{\partial t} + \left(-\sigma' t\right)\frac{\partial a}{\partial t} = 0,$$

which can be rearranged to solve for t. This yields the following expression for the second-best value of the tax, denoted t^*:

$$t^* = \frac{-\phi}{1-\sigma'\left(\frac{\partial a}{\partial t}/\frac{\partial e}{\partial t}\right)}. \tag{5}$$

Note that, when $\sigma' = 0$ as in the first-best, $t^* = -\phi$, which is again the Pigouvian tax. The second-best rate also equals the Pigouvian tax when the attribute a is perfectly inelastic ($\frac{\partial a}{\partial t} = 0$). When neither of those conditions holds, the second-best tax will not be the Pigouvian rate because the planner must trade off the distortion in the choice of a against the externality gain that follow from changing t.

In all of our applications, $\sigma' < 0$. Moreover, we expect that $\frac{\partial a}{\partial t}$ and $\frac{\partial a}{\partial t}$ will both be negative (consumers adjust to a higher tax—a smaller subsidy—both by lowering e and by lowering a). Thus, the existence of attribute-basing causes the second-best tax rate t to be *below* the Pigouvian rate because the denominator is greater than 1.

The difference between the first and second-best rates will be larger as the attribute slope gets steeper (as $|\sigma'|$ gets larger). The deviation is also larger when the response of the attribute a to policy is larger relative to the response of e. This can be seen better when rewriting equation 5 in terms of elasticities:

$$t^* = \frac{-\phi}{1-\epsilon_a^{\sigma}\frac{\epsilon_t^a}{\epsilon_t^e}}, \tag{6}$$

where ϵ_y^x is used to denote the elasticity of x with respect to y for several sets of variables.[9] Equation 6 shows that the second-best tax is driven further from the first-best rate when the elasticity of a is large compared to the elasticity of e. That is, the market can respond to the policy by changing both a and e. When the market response is tilted more heavily towards a, this will drive the second-best tax rate towards zero because the response in a is distortionary. When the response to a larger subsidy or tax is predominantly a response in e, the difference between the second-best

[9]The result follows from substitution based on the definition of the elasticities and algebraic simplification, which uses the fact that $e = \sigma(a)$ at the constrained optimum.

tax rate and the Pigouvian benchmark will be small.

The regulatory impact analysis of footprint-based CAFE in the U.S. assumed that footprint would not respond to the policy, in which case the second-best tax rate would be equal to the Pigouvian tax. In contrast, Whitefoot and Skerlos (2012) argue that the footprint will be responsive. In our empirical analysis, we estimate the response of both weight (a) and fuel economy (e) to the tightening of the Japanese standard (t) and find that the response in weight is significant and that the second-best tax rate is approximately 40% below the Pigouvian rate.

2.2.4 Mispricing of externality does not justify attribute basing

The analysis above focuses on the first-best allocation, in which attribute-basing has no role to play. But, might there be a second-best argument that makes attribute-basing useful? It is a standard intuition in optimal tax design that it is efficient to introduce a small tax in one market that is not distorted in order to reduce a pre-existing distortion in a second market because the initial distortion creates small welfare losses compared to marginal improvements in pre-existing distortions.

To see whether this intuition applies to attribute-basing, suppose that the optimal tax rate t is set at a rate too low ($|t| < \phi$), so that the e is below the social optimum and the policymaker cannot increase it (perhaps due to political constraints) by changing t. Faced with this choice problem, what is the optimal attribute slope?

We solve the linear case here, where $\sigma(a) = \sigma' \cdot a + \kappa$, so that $t(a, e) = t \cdot (e - \sigma' a + \kappa)$. In that case, the second-best attribute slope is characterized by solving the following problem:

$$\max_{\sigma'} U(a, e) - P(a, e) + \phi e.$$

The solution to this is:

$$\sigma'^{*} = \underbrace{\frac{t + \phi}{t}}_{-} \underbrace{\frac{\partial e / \partial \sigma'}{\partial a / \partial \sigma'}}_{-}. \tag{7}$$

Under the assumption that $|t| < \phi$, the first term is negative because $t < 0$. The second term is negative if a and e move in different directions in response to a change in the attribute slope σ'. In that case, which we expect is the most likely to prevail, the attribute slope will thus be *positive*,

which is opposite that of all observed policies discussed in this paper. If a and e both rise (or fall) in response to a change in the attribute slope, then σ'^* will be negative, but it will not have the slope of real world policies, which are fit to data (as we discuss further below). In short, a tax rate that is too small is unlikely to be a justification for the use of attribute basing.

There are many real world policies in which it appears that policymakers are constrained by political factors from setting regulations or taxes at the right level. It is tempting to think that the introduction of attribute-basing into regulations would give the policymakers greater flexibility and therefore help circumvent constraints on setting the right level. This is unlikely to be the case, however, because attribute-basing is a very unwieldy tool for affecting the choice of the product characteristic that creates the externality; it primarily acts as an implicit subsidy for the attribute itself, not the externality generating characteristic.

2.2.5 Distortions in a are first-order if a has an unpriced externality

In some real world attribute-based policies, the attribute upon which the target is based itself generates externalities. Take the example of Japanese fuel economy standards, in which case e is fuel economy and a is vehicle weight. Heavier vehicles pose greater risk to drivers of other vehicles in the case of a collision (Anderson and Auffhammer Forthcoming). This risk is unpriced (whereas the increased safety of those in the heavier car is a private benefit and therefore should be priced in the market) and represents a negative externality associated with weight.[10]

The model can accommodate this possibility by adding a second externality to the welfare function equal to $\omega \cdot a$, so $W = U(a, e) + x + \phi e - \omega a$, with $\omega > 0$. To see the implications of this for the distortion related to the use of attribute-based standards in the baseline model, we can consider an approximation of the welfare loss due to a movement of choices from the Pigouvian benchmark (the choice when $t = -\phi$ and $\sigma' = 0$), denoted (a^*, e^*), to that which is chosen by the consumer under the attribute-based policy, denoted (a', e'). The first-order Taylor expansion for

[10] In principle, the externality risk may be partly priced through insurance or legal liability. White (2004), however, argues that neither tort liability nor mandatory liability insurance prices safety externalities. In brief, tort liability requires negligence, not just that one be driving a dangerous vehicle. Liability insurance generally coves the cost of damages to a vehicle, but it is but a small fraction of the value of a life. In addition, rate differences across vehicles are very coarse and reflect average driver characteristics in concert with vehicle attributes.

this welfare loss, when a has an externality of ω, is:

$$\begin{aligned} W(a', e') - W(a^*, e^*) =& (a' - a^*) \left(\frac{\partial U(a^*, e^*)}{\partial a} - \frac{\partial P(a^*, e^*)}{\partial a} - \omega \right) \\ &+ (e' - e^*) \left(\frac{\partial U(a^*, e^*)}{\partial e} - \frac{\partial P(a^*, e^*)}{\partial e} + \phi \right) \\ =& - (a' - a^*)\omega. \end{aligned} \tag{8}$$

Note that $\frac{\partial U(a^*,e^*)}{\partial a} = \frac{\partial P(a^*,e^*)}{\partial a}$ and $\frac{\partial U(a^*,e^*)}{\partial e} = \frac{\partial P(a^*,e^*)}{\partial e} - \phi$, and several components therefore cancel. (This is just an application of the envelope theorem.) What remains is that, to a first-order approximation, the welfare loss from attribute-basing will be equal to the change in a times the externality associated with a.

Intuitively, this indicates that when the attribute a has an unpriced externality associated with it, the welfare loss from the introduction of a "small" attribute slope will be dominated by the externality associated with a. This is relevant for our empirical evaluation, where we use this reasoning to justify a focus on how Japanese fuel economy standards cause an increase in average vehicle weight.

Conversely, if the policymaker were aware of the externality associated with a, the attribute-based policy could be designed to correct both externalities by setting $t = -\phi$ and $\sigma' = \omega/\phi > 0$. That is, if there are two externalities, then the attribute-based standard can affect a Pigouvian tax on each. This would imply an attribute slope which has the opposite sign as what we observe in reality, suggesting that policymakers are not using the degree of freedom in the policy to correct the related externality.

2.2.6 Results are similar when $t(\cdot)$ has notches

In appendix B we explore the welfare implications of notched attribute-functions, that is, cases in which the tax function $t(\cdot)$ may be discrete. The Japanese fuel economy regulation, for example, has a target function that is stepwise in weight (a). The key results of that model is twofold. First, all of our intuitions about optimal policy design and distortions carry over directly to the case with notches. We relegate these results to the appendix because the insight is the same as our baseline. Second, in the case where the tax function is continuous in e but notched in a (like the Japanese

fuel economy regulation), all of the distortions come from products that chose to bunch at notch points in a. Products that do not choose a to be exactly at a notch are not distorted because the notched subsidies create only income effects, which do not affect the choice of a or e when utility is quasilinear. This is useful for interpreting our empirical results, and we return to these theoretical results when discussing the relevant empirical results below.

2.3 The model when there is no trading

Up to now we have focused on the case where a common tax rate applies to all consumers and firms, which corresponds to a regulatory policy that allows trading. This characterizes a number of regulations, including CAFE and feebates in the automobile market. Some attribute-based regulations apply to each and every product in a market, or they apply to each firm separately and trading is disallowed. Key examples are appliance product standards, which specify a minimum energy efficiency that all products are required to meet. Refrigerators in the U.S., for example, must meet a minimum energy efficiency that is a continuous function of its "adjusted volume" (equal to fresh volume plus 1.63 times freezer volume) and a discontinuous function of features, including door type (french or not), location of freezer (top, bottom or side) and whether there is through the door ice.

When trading is not allowed, products will face different shadow prices if each product is required to meet the same minimum. This is equivalent to a tax policy that imposes different tax rates on e across different products. This will cause the marginal costs of compliance to be different across goods, thereby implying inefficient provision of e. In this setting, attribute-based regulation may provide a benefit by (partially) equalizing shadow prices. This benefit can be weighed against the costs stemming from distortions in the choice of a.

To capture the nature of non-trading, it is more natural to work with an attribute based standard than a tax. And, in this case, it is necessary to consider heterogeneity. We model heterogeneity as a discrete set of types of consumers $n = 1, ..., N$, each of whom chooses a separate vehicle because they have a different flow utility function $U_n(\cdot)$. For notational simplicity, we assume that there are measure 1 total consumers, and each type has measure $1/N$. Given quasilinearity, it is inconsequential to assume that I is common across types, so we make that assumption to streamline notation.

Under a standard, the problem of the type n consumer is:

$$\max_{a,e} W = U_n(a,e) + [I + G - P(a,e)] - \lambda \cdot (e - \sigma(a)) + \phi \cdot E,$$

where λ is the shadow price on the regulatory requirement that $e > \sigma(a)$. This yields identical first-order conditions as the tax case above, where $t = \lambda$.

We assume that the planner solves a utilitarian social welfare function that puts equal weight on utility across types, maximizing with respect to the policy function $\sigma(a)$:

$$\max_{\sigma(a)} SWF = \sum_n \left(U_n(a_n, e_n) + [I + G - P(a_n, e_n)] + \phi \cdot e_n \right).$$

Note that in the case of a standard, the planner does not choose directly a shadow price (tax rate), but rather chooses one implicitly by setting the stringency of the standard.

2.3.1 How actual policies are drawn

To better understand the merits of attribute-basing in the case without trading, we first take note of how actual attribute-based policies were determined for automobiles in the U.S. and Japan. In the U.S., to determine the footprint-target function in the reformed CAFE, government officials used engineering information to estimate the footprint and fuel economy of a fleet of vehicles that used a suite of technological improvements deemed to be feasible in the near future. They then fit a line to the data to determine how quickly fuel economy declined as footprint rose. They used this estimated slope as the slope of the attribute-based target.

In Japan, the formulation is somewhat different, but it has similar implications. When considering a tightening of fuel economy standards, the Japanese government defines weight categories. Within each weight category it identifies the vehicle with the highest fuel economy. It then determines the required fuel economy rating for vehicles in that weight category in some future target year by multiplying the fuel economy of this "front runner" by some factor, requiring a percentage improvement over the leader in the current period.

By setting the target function to move smoothly through the distribution of a and e that emerges from the private optimum, a standard that forces each product to meet the standard

would equalize the average change in e required of products at each level of a. That is, denote the choice of e_n that would emerge in the absence of policy for consumer n, and denote by σ_n the standard applicable to that consumer. Let $\Delta e = \sigma_n - e_n$. Actual policies (apparently) seek to equalize Δe across consumers and products—i.e., it equalizes marginal *quantities*. An efficient policy equalizes marginal costs, not marginal quantities. There is a special case, however, in which equalizing marginal quantities is equivalent to equalizing marginal costs. We next describe that special case as a way of illustrating the inefficiencies of attribute-basing in the more general case.

2.3.2 A sketch of the second-best non-trading solution

We derive here some facts about the solution to the linear case, when $\sigma(a) = \sigma' a + \kappa$. To do so, we set up the planner's problem as maximizing the sum of the increase in the externality minus the distortion caused by changes in the choice of a and e in response to the policy.

To do so, we denote the distortion from the choice of e as a function $m(e - e_n)$ that is zero when $e = e_n$ and greater than zero and convex in e above e_n; m is the Harberger-triangle in the m "market". We define an analogous function l for the distortion from any mischance in a. The planner's problem can then be written as:

$$\max_{\sigma',\kappa} SWF = \sum_n \phi(e_n - \tilde{e}_n) - m(e_n - \tilde{e}_n) - l(a_n - \tilde{a}_n),$$

where $\tilde{a}_n$ and $\tilde{e}_n$ denote the private optimum of e and a chosen by consumer n in the absence of policy, and a_n and e_n are the final chosen attributes in the presence of the policy. For expositional simplicity, but without the loss of any interesting features of the problem, we assume that there are the same number of each type of consumer (so we do not need to carry a weight term around the summations) and we assume that the regulation is binding for all types (if not, then some would simply contribute zero to the welfare function). If the policy is binding for each n, then the final choice of e each type will be $e_n = \sigma' a_n + \kappa$. Then, the problem can be rewritten as:

$$\max_{\sigma',\kappa} SWF = \sum_n \phi(\sigma' a_n + \kappa - \tilde{e}_n) - m(\sigma' a_n + \kappa - \tilde{e}_n) - l(a_n - \tilde{a}_n),$$

After rearrangement, the first-order condition of this problem for κ yields:

$$\sum_n m'_n = \phi \sum_n \left(1 + \sigma' \frac{\partial a_n}{\partial \kappa}\right)$$
$$\mathrm{E}[m'_n] = \phi \left(1 + \sigma' \mathrm{E}\left[\frac{\partial a_n}{\partial \kappa}\right]\right),$$

where we use m'_n to denote the derivative of the m function—that is, the marginal (distortion) cost—which is positive. The simplification makes use of the fact from the consumer's optimal choice problem that $l'_n = -\sigma' m'_n$. The second version of the optimality condition characterizes the function in terms of expectations (we could alternatively write this in terms of means, but the expectational notation is useful below in drawing a method of moments analogy).

This condition says that, at the optimum, the average marginal cost of increasing e should be set equal to the marginal externality ϕ times a factor. The factor is less than one, because $\frac{\partial a}{\partial \kappa} > 0$, assuming that $\sigma' < 0$. Thus, unless σ' is zero, or unless a does not respond to a tightening of the standard (perhaps because it is perfectly fixed, which we consider below), then the optimal policy will set marginal costs below the externality. That is, the overall stringency of the standard, chosen by raising or lowering κ, will be below the stringency that would be set when there were no attribute basing and is below the Pigouvian benchmark, which would be ϕ. The greater is σ' (in absolute magnitude), and the more responsive is a to a tightening of the standards, the greater will be the gap between the stringency of the Pigouvian tax and the second-best attribute-based regulation in the no-trading case. This is analogous to the second-best t in the trading case above.

Additional insights come from taking the first-order condition with respect to σ' and rearranging:

$$\sum_n m'_n a_n = \phi \sum_n \left(a_n + \sigma' \frac{\partial a_n}{\partial \sigma'}\right)$$
$$\mathrm{E}[m'_n a_n] = \phi \mathrm{E}[a_n] + \phi \sigma' \mathrm{E}\left[\frac{\partial a_n}{\partial \sigma'}\right].$$

This does not have an immediate interpretation that is helpful, but the implication of this condition can be seen more clearly by transforming it into a statement about the covariance of m' and a that will occur at the optimum. This is done by substituting the first-order condition for κ, in expectation operator notation, into the definition of the covariance of m' and a and then relating this to the

first-order condition for σ'. This yields:

$$\text{cov}(m'_n, a_n) = \phi\sigma' \left(\text{E}\left[\frac{\partial a_n}{\partial \sigma'}\right] - \text{E}\left[\frac{\partial a_n}{\partial \kappa}\right] \text{E}[a_n] \right).$$

At the second-best optimum, there will be a covariance between marginal costs and a. The term in parentheses can be signed. The covariance will be positive. That is, types with higher a_n will have higher marginal abatement costs (on average) than those with lower a_n. (Note, however, that this is not a closed form solution. Both of our optimum conditions have σ' in them.)

As discussed above, actual policy seems to have fitted a line to observed data. This is exactly what one would via OLS, as we discuss further in the next section below, if the loss function m' were quadratic and uniform across n and a were fixed. In that special case, $\frac{\partial a_n}{\partial \kappa} = \frac{\partial a_n}{\partial \sigma'} = 0$ so that the optimal covariance of m'_n and a_n is zero. Then, the equilibrium condition above is the method of moments condition for OLS, as it says that the covariance of the residuals (which are m' when the loss function is uniform across all types) and the "x-variable", which is a in our case. Note that what this does is attempt to equalize marginal costs of abatement across all values of a, which is familiar intuition from any externality-reducing public policy.

The fact that this is not the general solution shows that the second-best attribute slope will not simply fit data. Instead, it will induce a covariance between the marginal abatement cost and a. That is, it will take the slope from data fitting and reduce the slope. The intuition is that fitting the data is good for equalizing marginal costs, but the steeper is the slope, the greater is distortion in the a market. Thus, the planner will fit the data, but then reduce the slope somewhat, trading off reductions in distortions in a across all n types against the inefficiency of getting different m' values across different types.

2.3.3 The problem for which actual policies are the solution

Actual policies appear to equalize Δe. This will be a second-best policy in the special case where marginal cost functions are equal across all products and the attribute a is exogenous (it does not respond to policy incentives).

To see this, define the marginal cost of improving e as the difference between marginal willingness to pay and marginal cost. At e_n, by definition, this difference is zero, and it is increasing as e rises

above e_n. Define that function as $m(\Delta e)$. Suppose that $m(\cdot)$ is the same for every type n. Efficient allocation of increases in e will equalize marginal costs (m') across all n, which is equivalent to equating Δe across all n when the m function is common across n. This is the most generous version of a model in terms of justifying actual policy.

In this setting, the planner's problem can be recast as maximizing the difference between the externality gains from increasing e past the private solutions and the sum of marginal costs of increasing e. If the planner can choose a unique σ_n standard for each n, then the problem is:

$$\max_{\sigma_n} \sum_n (\phi(\sigma_n - e_n) - m(\sigma_n - e_n)).$$

First-order conditions will equalize marginal cost with ϕ for each n, that is $\phi = mu'$. This is straightforward if the planner can choose σ_n for each n.

When the policymaker has an attribute standard instead, however, then their problem is:

$$\max_{\sigma(a)} \sum_n (\phi(\sigma(a) - e_n) - m(\sigma(a) - e_n)).$$

In the general case where $\sigma(a)$ is nonlinear, $\sigma(a)$ can be chosen to be σ_n so long as every product with a value of a_n has the same e_n. To the extent that there is heterogeneity in the privately optimal e_n, conditional on a_n, then the attribute-based function cannot perfectly replicate the first-best schedule. For illustration, consider the case where $\sigma(a)$ is linear affine, so that $\sigma(a) = \sigma' a + \kappa$. Then the planner's problem is:

$$\max_{\sigma(a)} \sum_n (\phi(\sigma' a + \kappa - e_n) - m(\sigma' a + \kappa - e_n)). \tag{9}$$

This problem now becomes identical to fitting a line to data, where m is the loss function.

To see this most clearly, suppose that $m(\Delta e) = \Delta e^2$, so that adjustment costs are quadratic. Then, the objective function in expression 9 is identical to minimizing the sum of squared residuals. The solution is the Ordinary Least Squares coefficients for bivariate regression, with a shift in the intercept due to the ϕ terms; specifically, $\sigma' = \frac{cov(e_n, a_n)}{var(a_n)}$ and $\kappa = \phi/2 + \bar{e} - \sigma' \bar{a}$.[11]

[11] This results in an average value of Δe of $\phi/2$, which is the first-order condition for a quadratic fn e when m is quadratic.

This solution represents a second-best optimal attribute policy. If $m(\Delta e) = \Delta e^2$, then the optimal $\Delta e = \phi/2$ for every n. Any attribute-based policy will set one required value $\sigma(a)$ that applies to all n that share an a. Any variation in e conditional on a implies inefficiencies that result from different Δe values for different n. In the case where the attribute-based policy is linear affine, then the deviations of the mandated Δe from the optimal $\phi/2$ are equivalent to the residuals from the bivariate regression. By extension, the R^2 from this regression summarizes the inefficiency of the attribute-based policy, even in the special case in which it is the second-best policy.

If the R^2 of the regression of e_n on a_n is 1, then the linear attribute schedule fits the optimal schedule perfectly, and the policy matches the first-best. But, as the R^2 gets lower, it implies that the average deviation between the actual improvement required of each type, Δe_n, and the optimal amount of $\phi/2$ is getting larger, and therefore inefficiency is greater. In the case of Japan, this R^2 is around .75, suggesting that the attribute slope there captures about three-quarters of the variation in marginal costs (in the special case where all vehicles have equal marginal cost).

The analogy to OLS is, of course, specific to the case where the marginal cost function m is quadratic. With some other loss function, the second-best schedule will have a different fit and a goodness-of-fit statistic other than the R^2 will be the appropriate welfare metric. But the basic intuition will carry over to other loss functions. The key intuition, which is general, is that the efficiency of the attribute-based schedule is tied to the correlation between the optimal value σ_n and the attribute amount a_n. To the extent that a_n predicts σ_n exactly, a fully flexible attribute-based policy will be able to capture a larger share of the welfare gains. The additional consideration is that this derivation rests on the assumption that a is inelastic and does not respond to the subsidy. We discuss the importance of that assumption next.

2.3.4 Attribute-basing is inefficient compared to trading

The above derivation was the most favorable possible interpretation of attribute-basing from an efficiency point of view. Even in that case, when the attribute does not perfectly predict energy efficiency, attribute-basing will fail to fully equalize marginal costs of abatement. In reality, the attribute-based policy will perform even worse for two reasons. First, the optimal improvement for each good or consumer Δe_n will vary if the marginal cost $P(a, e)$ and marginal utility $U(a, e)$ curves have different slopes at the private optimum.

Second, the derivation above assumed that a was exogenous, but our empirical evidence below makes clear that this is not so. When a is endogenous, then the attribute-based schedule will induce distortions in the choice of a as described in our baseline model. These costs will be weighed against benefits from marginal cost equalization. It is important to note also that distortion in a is unavoidable in an attribute-based system, so long as a is a choice made by market participants rather than an immutable characteristic. There is no way to design an attribute-based policy to achieve the first-best, even if all marginal costs line up and the R^2 is one, unless the attribute slope is zero.

These deviations from the first-best represent the costs of using attribute-basing in lieu of an efficient policy that equalizes marginal costs across all sources of the externality, as would occur either in a regulation with full trading or in a Pigouvian tax. We are currently extending our analysis to cover a more general description of the second-best attribute-based policy in the non-trading case, which will be a part of later drafts of this paper. What is clear from the discussion here is that, in the non-trading case, attribute-based regulation can generate benefits from marginal cost equalization, but equalizing marginal costs via attribute-basing will inevitably create distortions in the choice of a, which may in turn distort the choice of e, and will invariably fail to fully equalize marginal costs in real world circumstances. In contrast, trading would equalize marginal costs and create efficiency. Moreover, a uniform *tax* on e (as opposed to a regulation) would achieve the first-best regardless of the structure of and dispersion in costs. Thus, even in markets where trading is difficult or costly, there is an efficient instrument available.

2.4 What explains the use of attribute-basing?

Given the arguments made above, one might ask why attribute-basing exists. One possibility is that policymakers do not fully understand the perverse incentives and inefficiencies of these policies because they had not been previously modeled. A second possibility is that policymakers are forced to design the best possible regulatory policy without trading, due to some political economy forces that constrain the available policy instruments. In that case, our results describe and estimate the costs of operating under those constraints when making policy.

A third possibility regards distribution. If policymakers wish to transfer surplus from the consumers and producers of goods with low a towards those with high a, an attribute-based regulation

will achieve that to some degree. For the case of the U.S., it is widely believed that the Detroit Three pushed heavily for attribute-basing because they produce heavier, less fuel-efficient vehicles. It is clear that the attribute-basing in CAFE makes compliance easier for these companies. If this was a desired policy goal, then the costs that we document here from distortionary responses and inefficiencies represent the costs of achieving this transfer. Note that heavier vehicles tend to be more expensive (particularly given that cars and light-trucks are regulated separately) and thus are generally purchased by wealthier consumers. As a result, it is likely that wealthier consumers, on average, benefit from the attributing-basing. A distributional justification for attribute-basing is thus more likely to be made on the producer side than the consumer.

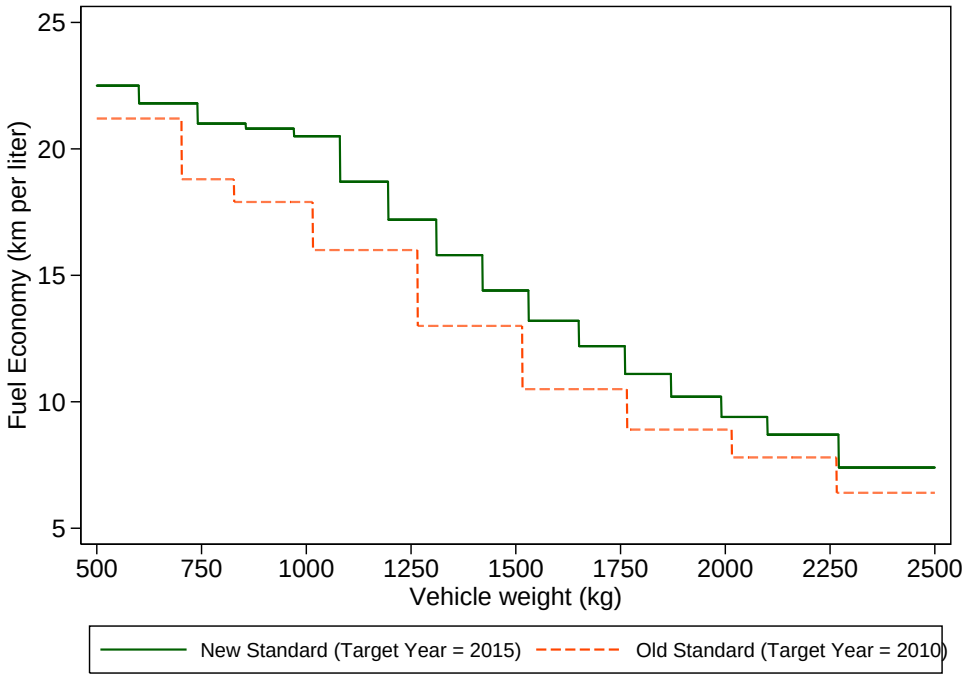
A fourth possibility is that policymakers may believe that attribute-basing is necessary to spur innovation. A firm can comply with a tightened regulation by downsizing their products (reduce a so that it is easier to raise e) or by investing in technology (to raise e holding a constant). In a static model, efficiency requires that all margins of adjustment be used until their marginal costs are equalized, but if the technology margin lowers production costs in the future period, then a dynamic model might suggest otherwise. In particular, if there are spillovers from technology, a policymaker might wish to increase the share of the adjustment due to technology beyond that which the market will choose on its own.

In appendix C we explore the possibility that attribute-basing is motivated by a belief that there are spillovers from technological advancement. In this case, the optimal choice of a policy instrument $t(e - \sigma(a))$ will involve attribute-basing, but this is an artifact of a setup that assumes both e and a have positive externalities attached to them. In that setting, attribute-basing provides a second instrument that can affect a Pigouvian tax on a. The optimal slope of the attribute function in this case does not resemble the attribute functions chosen by actual policymakers. Thus, while technological innovation is a plausible reason why some attribute-basing could be desirable, it does not appear to explain the policies that we actually see.

3 Empirical Analysis and Results

In this section, we investigate how the market responds to attributed-based regulation by analyzing the automobile market in Japan. The Japanese context provides several key advantages. First,

Figure 3: Fuel Economy Standard in Japan



Note: The dashed line shows the fuel economy standard in Japan until 2010. The solid line shows the new fuel economy standards whose target year is 2015.

the Japanese government has been using attribute-based fuel economy regulation since the 1970's. This gives us a long window for analysis and also allows us to use several policy changes for identification. Second, the Japanese fuel economy regulations feature notches in the attribute (weight) that determines the fuel economy target. These notches provide substantial variation in regulatory incentives and allow us to use empirical methods that are similar to the recent literature on nonlinear income tax schedules (Saez 1999 2010; Chetty, Friedman, Olsen, and Pistaferri 2011; Kleven and Waseem 2013). In contrast, the U.S. instituted attribute-based regulation for automobiles only in 2012, which leaves us little data for analysis, and its attribute-based target function is smooth, making identification more challenging.

3.1 Fuel Economy Standards in Japan

The Japanese government introduced the first attribute-based fuel economy standard in 1979. Since then, there have been four different schedules. Our data, which start in 2004, come from years

spanned by the two most recent regimes.[12] We summarize those two policies in Figure 3, which plots the fuel economy target for vehicles as a function of their weight. The downward slope in the figure shows that the required fuel economy is a decreasing function of weight. That is, heavier vehicles face a less stringent regulation. Furthermore, the regulation is notched. The declining function is not smooth, but rather is discontinuous at many weight cutoff points. Thus, for vehicles that have a weight near a cutoff point, marginal changes in weight lead to discrete changes in the fuel economy standard. This attribute-basing implicitly subsidizes vehicle weight, which we expect will lead to a distortion by leading to an increase in weight among firms attempting to meet the standard.

Automakers have two incentives to meet these fuel economy standards. First, they are obligated to have their sales-weighted average fuel economy exceed the minimum standard. Technically, this obligation extends to each weight segment separately, but firms are allowed to apply excess credits from one weight category to offset a deficit in another. Thus, in the end, the policy is functionally similar to the U.S. CAFE program, where there is one fleetwide requirement. Automakers have to pay fines if they cannot meet the standard by the target year of the regulation.

Second, in some years there are subsidies and tax exemptions that apply to individual cars if their fuel economy exceeds the standard for their weight class by a certain percentage. In 2009, the government introduced these subsidies as part of an economic stimulus package. Since then, the government has changed the eligibility requirements and the amount of the subsidies.

An important difference between the fuel economy regulation and the subsidy policy is that the later provides an incentive for an individual car to meet the fuel economy standard (as opposed to the fleet), and it thus creates a subsidy system that is notched in both fuel economy and weight. By contrast, the fuel economy regulation does not provide a direct incentive for an individual car to meet the target because what automakers have to do is to meet the target at the sales-weighted average. In the current version of this paper, we recognize that automakers have only the first incentive before 2009 and that they have both incentives after 2009. but we do not currently exploit the notches in the fuel economy dimension. We are currently developing strategies for exploiting the variation in subsidies and tax exemptions in recent years to conduct further empirical analysis.

[12] We are in the process of collecting data for the years before 2004, and we plan to extend our analysis when those data become available.

Table 1: Summary Statistics

Year	N	Fuel Economy (km/liter)	Weight (kg)	Displacement (liter)
2004	1,558	14.2	1,257	1.8
2005	1,529	14.3	1,252	1.8
2006	1,287	13.2	1,352	2.1
2007	1,289	13.3	1,368	2.1
2008	1,157	13.4	1,389	2.1
2009	1,230	13.4	1,393	2.2
2010	1,238	13.4	1,433	2.2
2011	1,156	13.6	1,444	2.2
2012	1,384	14.6	1,399	2.0
2013	1,488	14.1	1,484	2.3

Note: This table shows the number of observations, means of fuel economy, weight, and displacement for each year. Data are not sales weighted.

3.2 Data

We analyze data from the Japanese Ministry of Land, Infrastructure, Transportation, and Tourism (MLIT) that records fuel economy data for all new vehicles sold in Japan for each year between 2004 and 2013.[13] The record includes each vehicle's model year, model name, manufacturer, engine type, displacement, transmission type, weight, fuel economy, fuel economy target, estimated carbon dioxide emissions, number of passengers, wheel drive type, and devices used for improving fuel economy. The Ministry data does not include sales. Table 1 presents summary statistics of these data. There are between 1,100 and 1,600 different vehicle configurations sold in the Japanese automobile market each year. This includes both domestic and imported cars. The average fuel economy (not sales-weighted) declined slightly in 2006 and increased again around 2012. There is an overall positive trend in weight and displacement over time.

Although most of our analysis uses cross-sectional data, we do conduct panel data analysis in Section 3.5. To do so, we match data between different years in the following way. In our data set, each observation has a product identifier (ID). The product ID is a narrower definition than model name. For example, a Honda Civic may have several product IDs in the same year because there are Civics with different transmissions, displacements and drive types. We first match on product ID across years, which is often, but not always, constant over time. If automakers change the product

[13]The Ministry may have some data from before 2004. We are in the process of verifying availability of this earlier data.

ID between years, we match by using model name, displacement, drive type (e.g. four-wheel drive), and transmission (manual or automatic). That is, we consider two cars sold in two different years to be the same if they have matching IDs, or if they have the same model name, displacement, drive type, and transmission. We have experimented with matching on differing sets of characteristics and found that our key results are robust to the different criteria.

3.3 Excess Bunching at Notches in Fuel Economy Standard

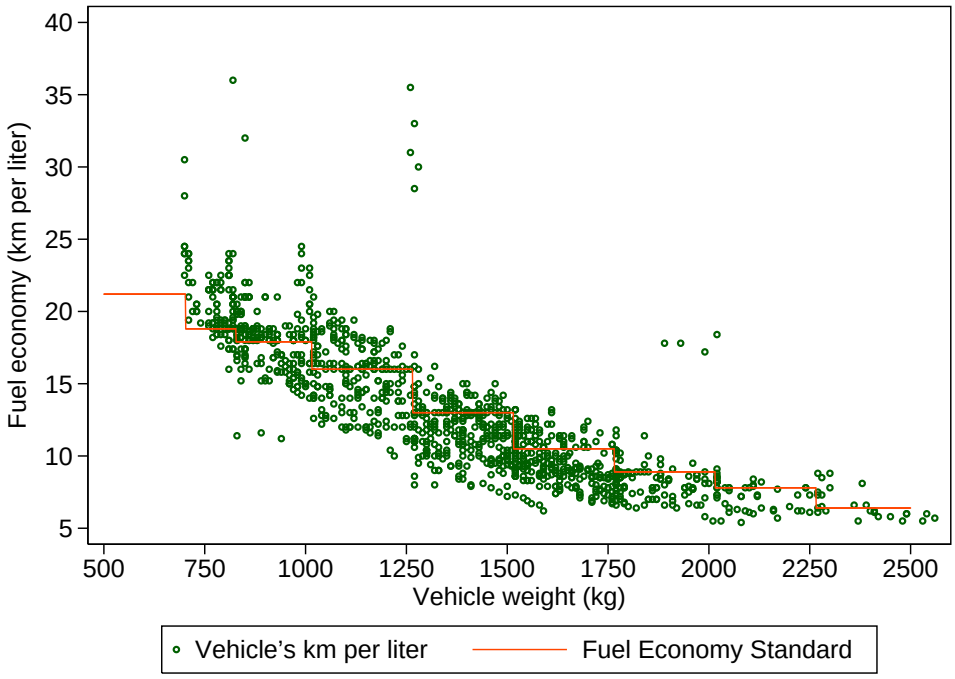
Our theoretical model indicates that the main welfare distortions resulting from attribute-based standards come from distortions in weight, and that for notched incentives, the distortion should be evident from excess mass in the weight distribution at cutoff values. We thus begin our analysis by looking at the distribution of the raw data to look for evidence of bunching. Figure 4 plots the fuel economy schedule for 2006, along with a scatter plot of each vehicle's fuel economy and weight in that year. We use 2006 as an example; the pictures for all other years reveal similar patterns.

The diagram reveals three things. First, there is a negative relationship between fuel economy and vehicle weight. The negative relationship comes from the fact that it is harder to achieve high fuel economy for heavier vehicles. Second, we observe "hugging" of many observations on the fuel economy standard schedule. That is, many vehicles have exactly the same fuel economy as the fuel economy standard, or just slightly higher fuel economy than the standard. Finally, and most prominently, there are excess numbers of observations at the notch points in vehicle weight, which occur at 710, 830, 1020, 1270, 1520, 1770, 2020, and 2270 kg.

To see the excess numbers of observations at the weight notches more clearly, we next provide histograms of vehicle weight superimposed on the notched schedules. Panel A of Figure 5 is the histogram of vehicles sold between 2004 and 2008. The bin size of the histogram is 10 kg. That is, each bar represents the density of the number of observations for each 10 kg bin. In this period, all vehicles had an old fuel economy standard schedule, which is presented in the same figure. There is visually clear excess bunching at each notch of the fuel economy standard schedule. The magnitude of the bunching is substantial. It is visually clear that the number of observations at each notch point is more than twice as large as the number of observations around the notch points.

Panel B of Figure 5 provides evidence that the excess bunching is the result of automakers' response to the notched schedule. After a policy change in fuel economy standard, automakers

Figure 4: Scatter Plot of Vehicle Weight (kg) against Fuel Economy (km/liter) in 2006



Note: This figure plots each vehicle's fuel economy (km per liter) in 2006 against the vehicle's weight. Each dot represents one observation.

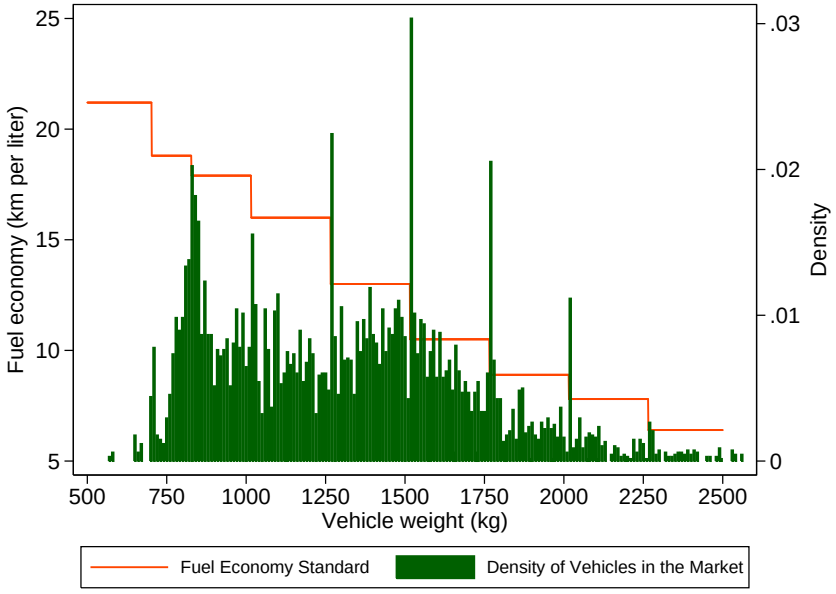
faced the new fuel economy standard schedule in 2013 as we present in Panel B. By comparing Panel A and B, one can see that the mass points shifted precisely in accordance with the change in the locations of the notch points. In sum, the raw data show strong evidence that the market responded to the attribute-based regulation by changing vehicle weight. Our theory suggests that, on its face, this is evidence of distortion. In the next section, we use an econometric method similar to that used in Saez (1999 2010); Chetty, Friedman, Olsen, and Pistaferri (2011); Kleven and Waseem (2013) to estimate the magnitude of the excess bunching and discuss the implications of this bunching for welfare.

3.4 Estimation of Excess Bunching at Notches

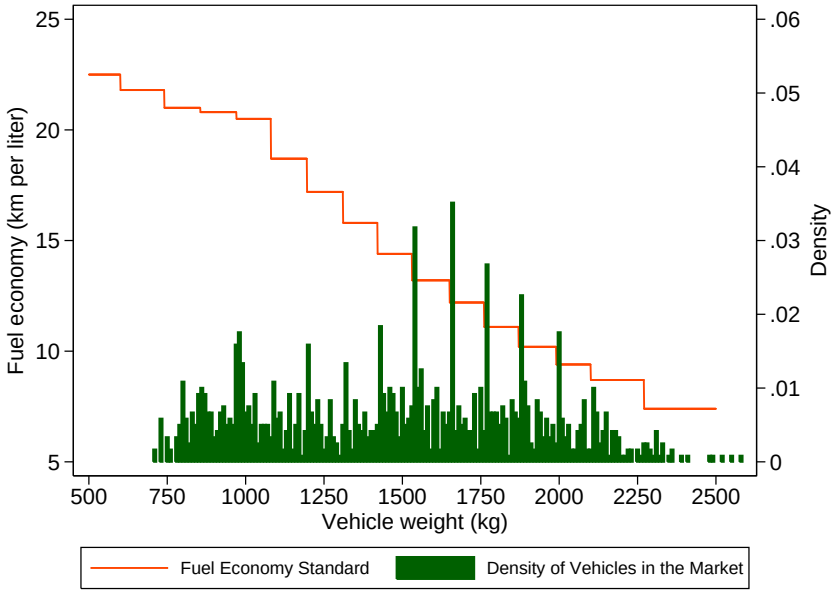
Econometric estimation of excess bunching in kinked or notched schedules is relatively new in the economics literature. Saez (1999) and Saez (2010) estimate the income elasticity of taxpayers in the U.S. with respect to income tax rates and EITC schedules by examining excess bunching around

Figure 5: Histogram of Vehicles and Fuel Economy Standard

Panel A. Years 2004 to 2008 (Old Fuel Economy Standard Schedule)



Panel B. Year 2013 (New Fuel Economy Standard Schedule)



Note: Panel A shows the histogram of vehicles from 2004 to 2008, where all vehicles had the old fuel economy standard. In 2013, essential all vehicles had the new fuel economy standard, which is presented in Panel B.

kinks in the U.S. personal income tax schedule. Similarly, Chetty, Friedman, Olsen, and Pistaferri (2011) estimates the income elasticity of taxpayers in Denmark with respect to income tax rates

by examining the excess bunching in the kinked tax schedules there. In Pakistan, the income tax schedule has notches instead of kinks. That is, the *average* income tax rate is piecewise linear. Kleven and Waseem (2013) uses a method similar to Chetty, Friedman, Olsen, and Pistaferri (2011) to estimate the elasticity of income with respect to income tax rates using bunching around these notches. Our approach is closely related to these papers, although our application is a fuel economy regulation, not an income tax.

To estimate the magnitude of the excess bunching, we first estimate the counterfactual distribution as if there were no bunching at the notch points. We start by grouping vehicles into small weight bins (10 kg bins in the application below) and denote by c_j the number of vehicles and by w_j the weight level in bin j. We fit a polynomial of order q to the bin counts in the empirical distribution, excluding bins in the range (w_L, w_U). We use $w_L = 10$, and $w_U = 10$ in the applications below. To minimize specification errors of the polynomial fit, we locally estimate the counterfactual distribution separately for each notch point. For each estimation, we include bins whose weight is larger than w_j - 200 and less than w_j + 200. For observations around each notch point k, we estimate:

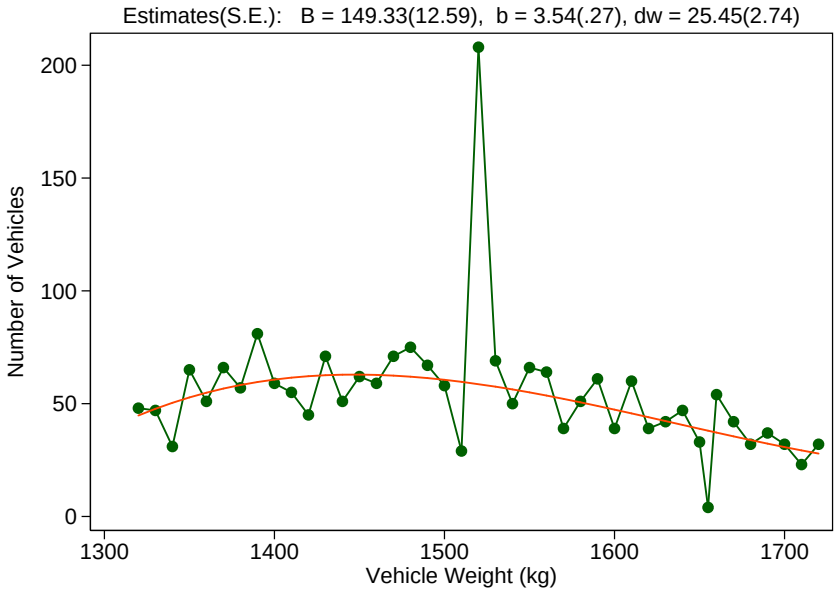
$$c_j = \sum_{p=0}^{q} \beta_p \cdot (w_j)^p + \sum_{h=w_L}^{w_U} \gamma_h \cdot 1\{w_j = h\} + \varepsilon_j, \tag{10}$$

where γ_h is a bin fixed effect for each bin in the excluded range (so that the regression gives a perfect fit in that range). The counterfactual distribution is then estimated as the predicted values from (10) omitting the contribution of the dummies in the excluded range, i.e. $\hat{c_j} = \sum_{p=0}^{q} \hat{\beta_p} \cdot (w_j)^p$. Excess bunching is the difference between the observed and counterfactual bin counts in the part of the excluded range at the notch point: $\hat{B} = \sum_{j=w_L}^{w_U} (c_j - \hat{c_j}) = \sum_{h=w_L}^{w_U} \hat{\gamma_h}$. We use $q = 3$ for the polynomial fit. Because we estimate the counterfactual distribution locally, using polynomials higher than third-order does not change our results. To provide precise estimates, we pool the data from 2004 to 2008 to estimate equation (10), although we still find fairly precise estimates of excess bunching even if we run the estimation separately for each year.

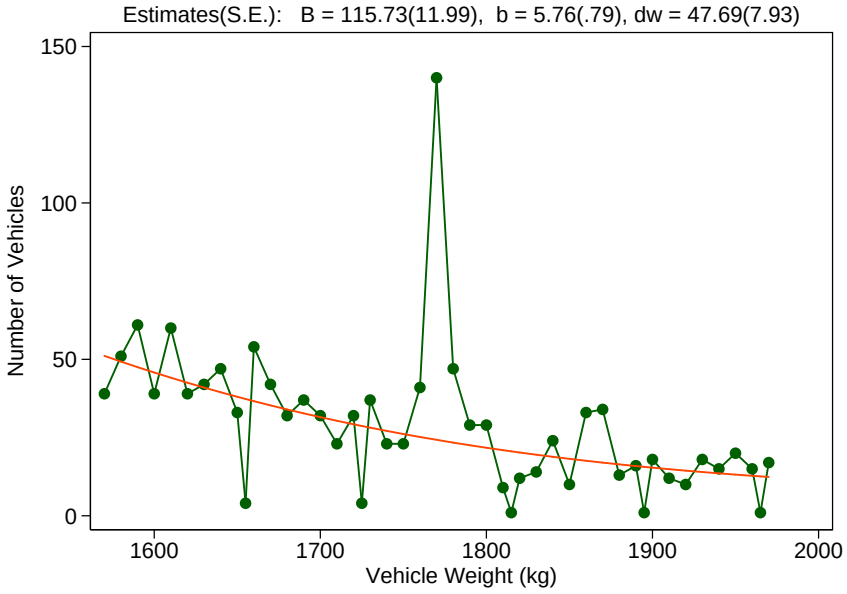
Figure 6 shows our estimation procedure graphically for two notch points. In Panel A, we plot the actual distribution and estimated counterfactual distribution around the 1520 kg notch point. Graphically, our estimate of excess bunching is the difference in height between the actual and counterfactual distribution at the notch point. The estimate and standard error of excess bunching

Figure 6: Graphical Illustration of Estimation of Excess Bunching at Each Notch Point

Panel A. Notch at 1520 kg



Panel B. Notch at 1770 kg



Note: This figure graphically shows the estimation in equation (10). The counterfactual distribution is estimated by the third-order polynomial of weight in the local area within 200 kg from each notch point. The figure also lists the estimates of B (excess bunching), b (proportional excess bunching), and dw (the range of weight affected by the notch). See the main text for details on these statistics.

Table 2: Estimation of Excess Bunching at Each Notch Point

Panel A. Notches at 710, 830, 1020, and 1270 kg

	(1)	(2)	(3)	(4)
Notches	710kg	830kg	1020kg	1270kg
Excess Bunching (B)	13.70	91.92***	58.21***	106.83***
(Excess number of vehicles at notches)	(28.63)	(28.32)	(17.70)	(13.53)
Proportional Excess Bunching (b)	1.34	2.99***	2.21***	3.31***
(Ratio relative to counterfactual dist.)	(0.74)	(0.67)	(0.4)	(0.36)
Δw (kg): Lower bound	3.48	19.94	12.18	23.13
Δw (kg): Uniform assumption	30.0	50.0	90.0	120.0
Δw (kg): Estimated from actual vs. counterfactual distribution	33.2	55.3	91.5	92.5

Panel B. Notches at 1520, 1770, 2020, and 2270 kg

	(5)	(6)	(7)	(8)
Notches	1520kg	1770kg	2020kg	2270kg
Excess Bunching (B)	149.33***	115.73***	64.59***	13.42***
(Excess number of vehicles at notches)	(12.59)	(12.00)	(7.01)	(3.91)
Proportional Excess Bunching (b)	3.54***	5.76***	6.66***	3.93***
(Ratio relative to counterfactual dist.)	(0.27)	(0.79)	(1.08)	(1.15)
Δw (kg): Lower bound	25.45	47.69	56.6	29.3
Δw (kg): Uniform assumption	120.0	117.0	120.9	122.4
Δw (kg): Estimated from actual vs. counterfactual distribution	149.1	94.5	130.6	121.8

Note: This table shows the regression result in equation 10. The dependent variable is the number of vehicles in each of 10 kg bins. We include a constant term and first-, second-, and third-order polynomial of weight in the regression. Standard errors are in the parentheses. ***, **, and * show 1, 5, and 10% levels of statistical significance from zero.

B is 149.33 (12.59). We also define the proportional excess bunching, b, which is the ratio between the actual and counterfactual distribution at the notch point. In this case, $b = 3.54$ (0.27), which means that the observed distribution has 3.54 times more observations than the counterfactual distribution at the notch point. Similarly, we illustrate our estimation around the 1770 kg notch

point. At this notch, $B = 115.73$ (11.99) and $b = 5.76$ (0.79), meaning that the height of the excess bunching is 5.76 times higher than the counterfactual distribution at the notch point.

Table 2 presents our estimation results for each notch. Excess bunching is statistically significant from zero for all notches except for the first notch point, around which there are very few observations (see Panel A of Figure 5). The results show that proportional excess bunching, b, is particularly large for the notch points for heavier vehicles.

What do the estimates of B and b tell us about welfare? Our theoretical model indicated that, under our assumptions about market structure, excess bunching should come from the "left"; that is, automakers should increase vehicle weight to reach notch points, not decrease it. Based on that assumption, we describe several different approaches to estimate the increase in weight, Δw.

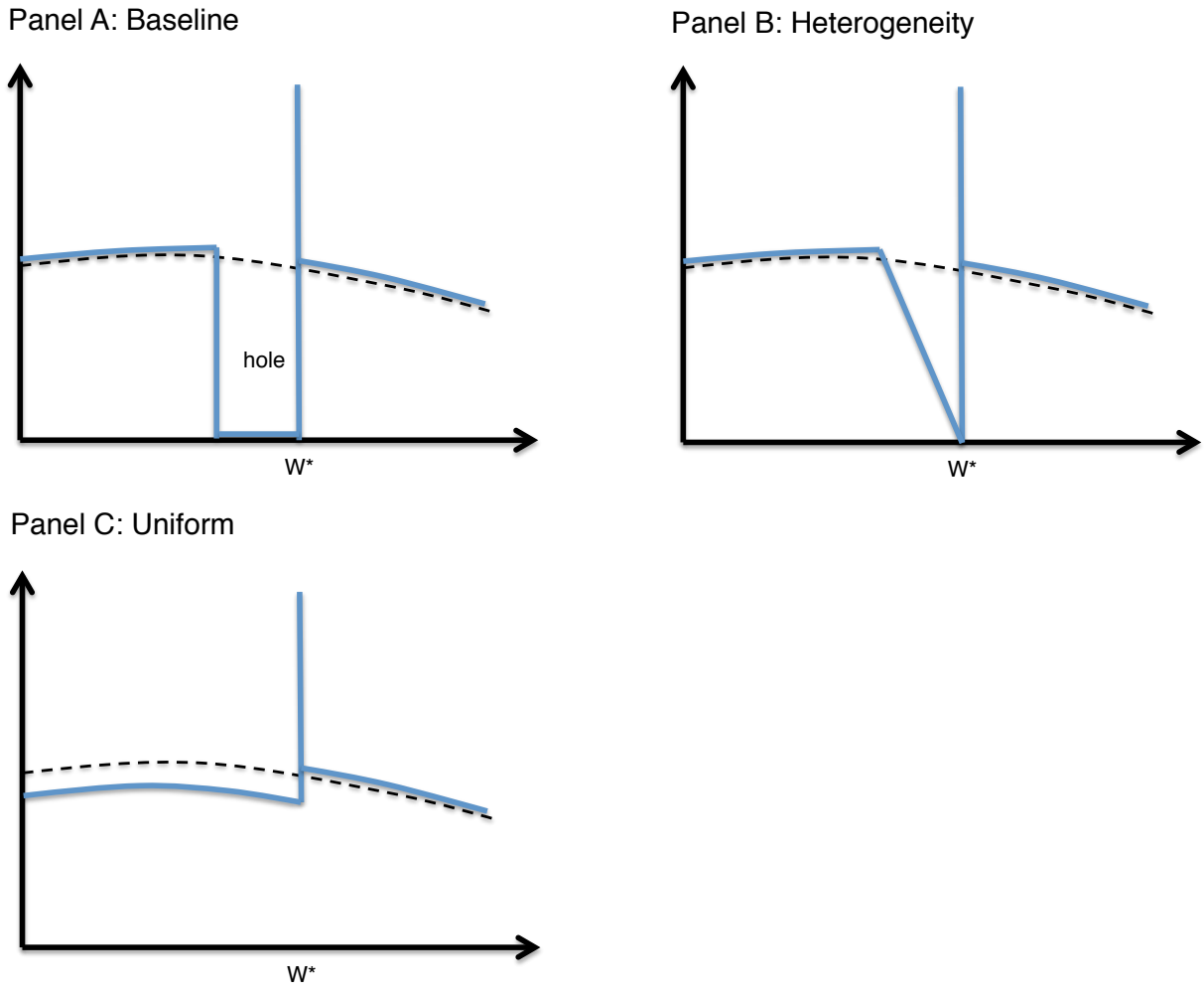
First, we begin with a baseline case where there is no heterogeneity in the elasticity of weight with respect to fuel economy standards. This case is conceptually equivalent to the baseline case in Kleven and Waseem (2013). All firms have a common elasticity of weight with respect to fuel economy standards. In this case, we can use the estimates of B and b to estimate how much automakers increase vehicle weight in response to the attribute-based fuel economy regulation. Consider a notch point at $w*$. Suppose that vehicles with weight $\in (w^* - \Delta w, w^*)$ bunch at w^* because of the notch schedule at w^*. Denoting the counterfactual distribution of vehicle weight by $h_0(w)$, excess bunching B at notch point w can be characterized by:

$$B = \int_{w^*-\Delta w}^{w^*} h_0(w)dw \approx h_0(w^*) \cdot \Delta w^*. \tag{11}$$

The approximation comes from the fact that the counterfactual distribution is fairly close to a uniform distribution at the local area of each notch point. Because our estimation provides estimates of B and $h_0(w^*)$, we can calculate estimates of Δw^* as $\Delta w^* = B/h_0(w^*)$. This standard case is described in Panel A in FIgure 7. Because all vehicles within a certain range of weight will bunch at a notch point, there should be a complete hole on the left hand side of the notch. However, this is not consistent with our data; there is no complete hole in each notch point.

The second approach in Kleven and Waseem (2013) provides a potential explanation for non-existence of holes, by considering heterogeneity in elasticity. Suppose that the elasticities of weight with respect to fuel economy standards are distributed between 0 and $\bar{e}$. At each elasticity level,

Figure 7: Density Distributions Under Different Assumptions



behavioral responses to the notch can be characterized as in the baseline case. The bunching segment for those with elasticity e is given by $(w^*, w^* + \Delta w_e^*)$, where Δz_e^* is increasing in e. The post-notch weight density will look like the solid blue curve in Panel B. In the case of heterogeneity, bunching can be used to estimate the average weight response $E[\Delta w]$. Denoting by $h_0(w, e)$ the joint weight-elasticity distribution in the baseline and by $h_0(w) \equiv \int_e h_0(w, e)de$ the the unconditional weight distribution, we have

$$B = \int_e \int_{w^* - \Delta w_e^*}^{w^*} h_0(w, e)dwde \approx h_0(w^*) \cdot E[\Delta w_e^*]. \tag{12}$$

When there is heterogeneity in elasticity, there will no more complete hole as described in Panel B.

Still, there should be a small hole in ranges close to the notch point. In addition, the empty areas should be shrinking as the weight deviates from the notch. However, Figure 6 does not suggest such evidence in our data; the empirical distribution is quite different from Panel B of Figure 7. The empirical distribution appears to be closer to a uniform distribution in each weight category.

The third approach is to assume that bunching comes uniformly from a lower weight category. This appears to be more consistent with the resulting shape of the empirical distributions. In this case, $E[\Delta w^*]$ is equal to the half of the difference in weight between the notch point and a lower notch point. This approach provides larger values for $E[\Delta w^*]$ compared to the last two approaches. However, the estimate based on this approach can be still underestimated if bunching comes from further lower weight categories.

Finally, as the fourth approach, we empirically estimate where bunching comes from. We compare the counterfactual density and actual density and calculate the residuals between the two densities for each weight points on the left hand side of the notch. Then, using the residuals, we calculate the weighted average of Δw^*. In our data, the residuals become close to a uniform distribution in most weight categories. As a result, the results based on this approach become very close to the results with the uniform assumption.

The last three rows of Table 2 show Δw^* for the second, third, and fourth approaches. Although the second approach is clearly inconsistent with our data, we want to keep the results as our lower bounds of Δw^*. We consider that Δw^* based on the actual and counterfactual distribution is more consistent with our data. Therefore, we describe these results as our main results. As Figure 6 suggests, Δw^* based on this method becomes very close to the case with the uniform assumption.

With the uniform assumption and the last approach, the estimates of Δw^* range from 30 kg to 130 kg. In addition, our lower bound estimates of Δw^* range from 3 kg to 56 kg. In the welfare analysis section, we provide welfare measures based on the estimate of Δw^* from each approach.

Overall, the results in this section provides evidence that automakers significantly respond to the attribute-based fuel economy regulation by changing their vehicle weight. We show that given the assumption that bunching comes from the left (lower weight) of each notch, our estimates of excess bunching imply that automakers increase vehicle weight between 6 kg and 28 kg to reach less stringent fuel economy regulations. In the next section, we exploit our panel data and a policy change in fuel economy regulation to test the assumption about the direction of bunching.

Table 3: Changes in Vehicle Weight from 2008 to 2013: Does Bunching Come from Left or Right?

	(1) ΔWeight	(2) ΔWeight	(3) ΔWeight	(4) ΔWeight
1{Bunching in 2013}	13.79***	13.23***	13.62***	12.48**
	(5.25)	(5.04)	(4.98)	(5.07)
(Weight in 2013)/1000		46.14***	261.34***	-131.18
		(6.83)	(54.79)	(346.21)
(Weight in 2013)/1000)^2			-71.47***	185.45
			(18.06)	(224.48)
((Weight in 2013)/1000)^3				-54.03
				(47.05)
Constant	24.56***	-42.32***	-196.26***	-4.05
	(2.81)	(10.26)	(40.19)	(172.16)
Observations	531	531	531	531
R^2	0.013	0.091	0.118	0.120

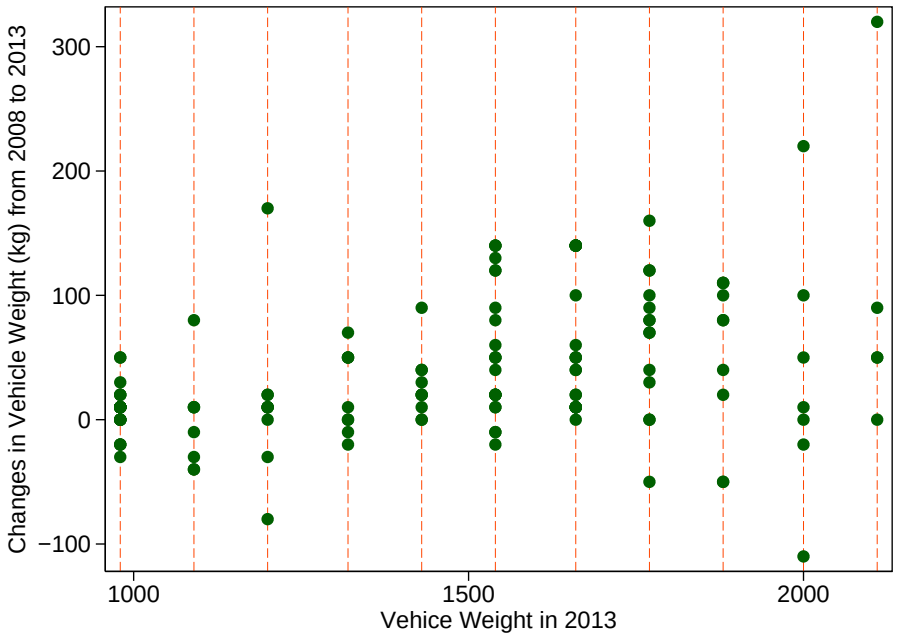
Note: This table shows the regression result in equation 13. The dependent variable is the change in vehicle weight (kg) from 2008 to 2011. 1{Bunching in 2013} is a dummy variable or samples that bunch in the fuel economy standard schedule in 2013. Standard errors are in the parentheses. ***, **, and * show 1, 5, and 10% levels of statistical significance from zero.

3.5 Panel Data Analysis: Where does bunching come from?

The previous section provided visually clear and statistically significant evidence of bunching at the notch points in the weight dimension. Theory suggests that bunching should come from the left (weight should increase), but we wish to test this assumption with the data. To do so, we exploit a policy change in the Japanese fuel economy standard that changed the location of the notched schedules between 2008 and 2013.

The two policy regimes are shown in Figure 3. Before 2008, automakers faced the old fuel economy standard represented by the dashed line. In 2009, the new standard was introduced. The years between 2009 and 2012 were a transition period, during which time automakers were allowed to choose which schedule they wished to use. In 2013, all automakers were required to be on the new standard. We focus our comparison on 2008 and 2013 because the automakers did not have a choice about the schedule in those years. For these two years, we use our panel data, which links the same car type over time, to see how vehicle weight and fuel economy changed.

Figure 8: Changes in Vehicle Weight from 2008 to 2013 for Samples that Bunch in 2013

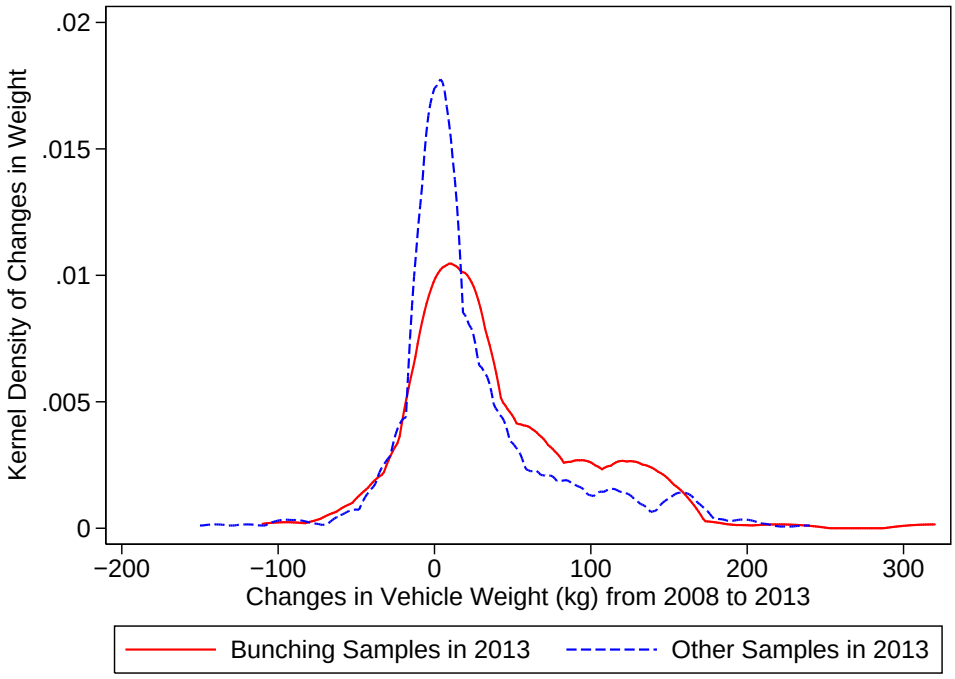


Note: This figure plots the raw data of individual vehicles that bunch in the notch points in the fuel economy standard in 2013. The horizontal axis shows the vehicle weight in 2013, the vertical axis shows the change in weight from 2008 to 2013, and the dashed vertical lines present the notch points.

We take two approaches to investigating whether bunching comes from the left or right. Our first approach is to examine whether vehicles that we observed bunching in 2013 featured an increase or a decrease in their weight between 2008 and 2013. In Panel B of Figure 5, we see significant bunching within 10 kg of the notches at 970, 1090, 1200, 1320, 1430, 1540, 1660, 1770, 1880, 2000, and 2110 kg. In Figure 8, we take the sample of those vehicles and plot their change in weight between 2008 and 2013 against their final weight in 2013. The horizontal axis shows the vehicle weight in 2013, the vertical axis shows the change in weight from 2008 to 2013, and the dashed vertical lines show where there are weight notches. For most vehicles in the diagram, the change in weight is positive, which suggests that bunching comes from weight increases.

There was an overall increase in weight, however, in vehicles between those years, so that Figure 8 may be partly the product of a secular trend. To account for any secular trend in weight, we use the vehicles that do not bunch in 2013 as a control group to establish a counterfactual. We focus on vehicles that have a weight between 960 and 2020 kg, which represents a significant majority of

Figure 9: Kernel Density of Changes in Vehicle Weight from 2008 to 2013



Note: This figure shows the kernel density of the changes in weight from 2008 to 2013 for the bunching samples and other samples in 2013.

the market, as a control group because light vehicles (those weighing less than 800 kg) appear to have had a significantly different weight trend.[14]

Figure 9 shows the kernel density of the changes in weight from 2008 to 2013 for the vehicles that bunched in 2013 and those that did not. The sample of those who bunched is pushed towards the right relative to the control group of non-bunchers. In particular, the distribution indicates that there are many vehicles who bunched at a weight notch after increasing weight by 50 to 150 kg.

To estimate the difference in the change in weight between the bunching and non-bunching vehicles, we estimate the following equation by OLS:

$$\Delta weight_i = \alpha + \beta \cdot 1\{bunching_i\} + f(weight_{i,2013}) + \varepsilon_i, \tag{13}$$

[14]Note that adding the excluded light cars does not change results in our regression analysis, below, in which we include a polynomial that allows the trend to differ vehicles of different weights.

where $\Delta weight_i$ is the change in weight (kg) from 2008 to 2013 for vehicle i and $1\{bunching_i\}$ equals one if vehicle i is located within 10 kg of a weight notch in 2013. Regressing $\Delta weight_i$ on $1\{bunching_i\}$ estimates a difference in mean weight between the two groups. This simple regression potentially produces a bias if the change in weight is systematically larger for lighter or heavier vehicles. In our case, any potential bias is likely small because the bunching locations are widely spread in the weight dimension so that the bunching and non-bunching samples have fairly similar weight on average. Nevertheless, we aim to eliminate any bias that might exist by controlling for a polynomial in $weight_{i,2013}$.

Table 3 presents regression results. Column 1 shows that the simple mean difference in the change in weight between the bunching samples and others is 13.79 kg. That is, vehicles that bunched increased their weight by 13.79 kg more than other vehicles. In column 2, we control for any linear relationship between the change in weight and final weight by adding a linear control for weight in 2013. As expected, this yields a positive coefficient on weight, which means that heavier vehicles had slightly larger increases in weight, but this has very little impact on our coefficient of interest, as expected. We include the second- and third-order polynomials in columns 3 and 4, which has no significant impact on our coefficient of interest.

These regressions and figures indicate that, on average, bunching comes from the left of the notch points. That is, automakers *increase* their vehicle weight in response to the attribute-based regulation notches, and our estimates suggest that the mean of this weight increase is about 13 kg. Note that this is smaller than our estimate in the previous section, which is estimated by bunching around notches in the old fuel economy standard schedule. The new fuel economy schedule has smaller steps compared to the steps in the old fuel economy standard schedule. Therefore, the new standard may create smaller incentives to change weight. We go further in the next section by using variation in incentives created by the policy change to estimate a weight response elasticity.

3.6 Panel Data Analysis: Elasticity of Attributes with Respect to Regulation

Our theory revealed that if a planner were to choose the second-best tax rate on fuel economy, taking as given the slope of the attribute-based target, that the deviation between the second-best tax rate and the Pigouvian benchmark would be a function of the relative elasticities of weight and fuel economy when the policy was tightened. In this section we estimate those elasticities and

provide a back of the envelope calculation as to how much the attribute-basing would lower the optimal (second-best) stringency of the policy.

Figure 3 shows that the new regulation increased the fuel economy requirement for vehicles in all weight classes, but the increase was of a different magnitude for vehicles in different weight classes. For example, compare a vehicle that weighed 1420 kg in 2008 to one that weighed 1430 kg in that year. The former vehicle faced an increase in its fuel economy standard from 13 to 15.8 km/liter (a 22% increase), while the latter vehicle faced an increase in its fuel economy standard from 13 to 14.3 km/liter (an 11% increase).[15] In this section, we exploit the discontinuities in the variation in the changes in fuel economy standards to estimate the elasticity of attributes with respect to fuel economy regulation.

We use a regression discontinuity design (RDD) that is similar to the identification strategy in Saez (2003). Saez uses discontinuous variation in income tax rates created by "bracket creep" in the U.S. income tax schedule. In his study, there is no policy change in the income tax schedule in nominal terms. However, inflation is high enough to shift the income tax schedule towards the right in real terms, causing taxpayers to face differing real income tax schedules. In our context, we have changes in 1) the locations of the kink points and 2) the stringency of the fuel economy regulation, which provides powerful variation enabling us to estimate how variation in the stringency of regulation affects attribute changes.

Our estimation is based on the following OLS regression:

$$\Delta weight_i = \alpha + \beta \cdot \Delta standard_i + \gamma \cdot X_i + \varepsilon_i, \tag{14}$$

where $\Delta weight_i$ is the change in weight (kg) from 2008 to 2013 for vehicle i, $\Delta standard_i$ is the change in the fuel economy standard (km/liter) for i, and X_i is a vector of control variables. OLS estimates will be biased because $standard_{it}$ is a function of $weight_{it}$ for $t \in (2008, 2013)$. If there are unobservable shocks in ε_i, they will affect $\Delta standard_i$ as well as $\Delta weight_i$. Because the fuel economy standard is a decreasing step function of weight, we expect that OLS estimation produces a downward bias for β.

To address the endogeneity, we use a policy-induced change in the standard as an instru-

[15]Note that this variation is driven by the front-runner in each category, which we treat as plausibly random variation.

Table 4: Elasticity of Attributes (weight) with Respect to Fuel Economy Standards

	ΔWeight(kg)			Δln(Weight)	
	(1)	(2)	(3)	(4)	(5)
ΔFuel Economy Standard	-8.67***	18.70**	18.86***		
	(1.69)	(7.37)	(7.26)		
Δln(Fuel Economy Standard)				0.16***	0.16***
				(0.05)	(0.05)
(Weight in 2008)/1000		479.58***	313.30	0.24***	0.18
		(76.40)	(434.38)	(0.04)	(0.26)
(Weight in 2008)/1000)^2		-146.39***	-36.23	-0.08***	-0.04
		(24.04)	(278.79)	(0.01)	(0.17)
((Weight in 2008)/1000)^3			-23.40		-0.01
			(58.19)		(0.04)
Constant	42.00***	-371.45***	-291.45	-0.18***	-0.14
	(3.51)	(66.05)	(224.59)	(0.03)	(0.13)
Observations	531	531	531	531	531
Estimation	OLS	RD(2SLS)	RD(2SLS)	RD(2SLS)	RD(2SLS)

Note: This table shows the regression result in equation 14. The dependent variable is the change in vehicle weight (kg) from 2008 to 2011 or log of the change in weight. Δ Fuel Economy Standard is the change in fuel economy standard. Standard errors are in the parentheses. ***, **, and * show 1, 5, and 10% levels of statistical significance from zero.

ment, as in Saez (2003); Saez, Slemrod, and Giertz (2012); Ito (Forthcoming). Specifically, we use $\Delta standard_i^{PI} = standard_{i,2013}(weight_{i,2008}) - standard_{i,2008}(weight_{i,2008})$. This instrument, which is sometimes called a simulated instrument, computes the predicted change in the standard at a weight level $weight_{i,2008}$. The instrument thus captures the change induced by the policy change for the weight level $weight_{i,2008}$. To be a valid instrument, $weight_{i,2008}$ has to be uncorrelated with ε_i, because the instrument is a function of $weight_{i,2008}$. This condition can be violated if lighter cars and heavier cars in 2008 have different underlying changes in weight. The advantage of our identification strategy is that we can include any smooth controls for $weight_{i,2008}$ in X_i to account for such concerns. Because the instrument's variation in the change in standard is discontinuous in weight, including any smooth controls for $weight_{i,2008}$ does not eliminate our ability to identify the coefficient. In our estimation, we use $\Delta standard_i^{PI}$ as an instrument for $\Delta standard_i$ and include

Table 5: Elasticity of Attributes (fuel economy) with Respect to Fuel Economy Standards

	ΔFuel Economy (km/liter)			Δln(Fuel Economy)	
	(1)	(2)	(3)	(4)	(5)
ΔFuel Economy Standard	-0.00	0.22	0.20		
	(0.06)	(0.21)	(0.21)		
Δln(Fuel Economy Standard)				0.22	0.21
				(0.16)	(0.16)
(Weight in 2008)/1000		7.80***	23.07*	0.56***	1.25
		(2.18)	(12.34)	(0.13)	(0.79)
(Weight in 2008)/1000)^2		-2.61***	-12.73	-0.19***	-0.65
		(0.69)	(7.92)	(0.04)	(0.51)
((Weight in 2008)/1000)^3			2.15		0.10
			(1.65)		(0.11)
Constant	0.26**	-5.61***	-12.95**	-0.40***	-0.73*
	(0.12)	(1.89)	(6.38)	(0.10)	(0.39)
Observations	531	531	531	531	531
Estimation	OLS	RD(2SLS)	RD(2SLS)	RD(2SLS)	RD(2SLS)

Note: This table shows the regression result in equation 14 but uses $\Delta FuelEconomy_i$ as the left-hand side variable. $\Delta FuelEconomy_i$ is the change in fuel economy (km/liter) for vehicle i from 2008 to 2013. Standard errors are in the parentheses. ***, **, and * show 1, 5, and 10% levels of statistical significance from zero.

a third-order polynomial of $weight_{i,2008}$ in X_i.[16]

Table 4 presents estimation results. As expected, the OLS estimate in column 1 is lower than our RD estimates in columns 2 and 3. This is likely due to the downward bias stemming from the endogeneity of $\Delta standard_i$. In columns 2 and 3, we show our RD estimates. The first stages are very strong because the policy-induced change in fuel economy standard strongly affects the actual change in standard. After we include the first- and second-order of polynomials in $weight_{i,2008}$, our estimates are not sensitive to the inclusion of higher-orders of polynomials. The estimates imply that a one unit increase in the fuel economy standard (km/liter) results in an increase in vehicle weight of19 kg. In columns 4 and 5, we use a log-log RD specification to estimate a constant elasticity. The estimate implies that an 1% increase in the fuel economy standard results in a 0.16% increase in vehicle weight. These results provide empirical evidence that automakers change their vehicle weight in response to the change in the stringency of fuel economy regulation.

[16]Including yet higher orders of the polynomial does not change our results.

In Table 5, we estimate equation 14 but use $\Delta FuelEconomy_i$ as the left-hand side variable. $\Delta FuelEconomy_i$ is the change in fuel economy (km/liter) for vehicle i from 2008 to 2013. In this estimation, we test whether increases in the fuel economy standard cause an increase in the fuel economy of vehicles. The estimates indicate that more stringent fuel economy standards result in increases in fuel economy, but the estimates are not statistically significantly different from zero.

We can use equation 6 to translate these elasticity estimates into a policy implication. The formula relates the second-best tax rate (or shadow price) t to the externality ϕ and the elasticities of a and e with respect to the policy, and the elasticity of the target fuel economy rating with respect to weight:

$$t^* = \frac{-\phi}{1 - \epsilon_a^\sigma \frac{\epsilon_t^a}{\epsilon_t^e}} = \frac{-\phi}{1 - (-.89)\frac{.16}{.21}} = .60(-\phi).$$

We estimate the elasticity of the policy function by regressing the log of the fuel economy values for each step on the log of the weight, measured at the midpoint, of the steps for the 2008 policy and get an elasticity of -.89. We then plug in our estimates of the elasticities of weight and fuel economy with respect to a tightening of the policy from tables 4 and 5. Together, these estimates imply that the second-best tax rate would be equal to marginal damages divided by 1.68, which is equivalent to saying that the second-best tax rate is 60% of the Pigouvian tax. That is, because weight responses strongly to increases in the standard, the policy-maker would choose a tax rate significantly below marginal damages from the externality in order to balance the distortions from the choice of a against the externality improvement.[17]

In summary, our empirical analysis provides three findings: 1) there is clear and statistically significant evidence that vehicles bunch at the attribute (weight) notch points in the Japanese fuel economy standard; 2) given the assumption (backed by theory) that bunching comes from the left (i.e., bunching comes from increases in weight), the excess bunching in our cross-sectional variation implies that the average increase in vehicle weight for each notch point is between 6 kg and 28 kg; and 3) our panel data analysis provides supporting evidence for this assumption, as we find that a one unit increase in the fuel economy standard results in an increase in vehicle weight of 19 kg and a one percent increase in fuel economy standard results in an increase in vehicle weight by 0.16%.

[17]Note that our estimate of the fuel economy elasticity is imprecise, so that the standard error around this statistic is large, though even the largest values of the fuel economy elasticity still imply significant deviations from Pigou. For example, if we take the upper value of the elasticity's 95% confidence interval, but continue to use the point estimate for the weight elasticity, the reduction in the second-best tax relative to the Pigouvian benchmark would be 21%.

3.7 Welfare implications of weight manipulation

As emphasized in the theory section, the main welfare loss from the response to attribute based standards in the situation where the attribute itself has an unpriced externality will be the inefficiency from an increase in the attribute. An approximation of the distortion from the policy can therefore be calculated by multiplying the change in the attribute by the externality, as indicated in equation 8. Our empirical estimates above suggest that the increase in weight caused by the attribute-basing in the Japanese car policy is between 30 and 130 kg based on our bunching estimation and also an alternative method provide estimates between 10 and 20 kg as lower bounds. The weighted average weight increase across all cars that bunch implied by the bunching methodology in table 2 is 104 kg and the lower bound estimate is 15 kg. We provide our main welfare implications based on 104 kg and also provide the same calculation based on our lower bound estimate of the change in weight, which equals to 15 kg.

To estimate the externality associated with increased weight, we use estimates from Anderson and Auffhammer (Forthcoming), which concludes that an increase in vehicle weight of 1000 pounds (454 kg) is associated with a 0.09 percentage point increase in the probability that the vehicle is associated with a fatality, on a mean probability of 0.19%. We use a standard estimate of the value of a statistical life (VSL) of $7 million. (Both of these estimates come from the U.S., which is a potential weakness of the calculation.)

We thus calculate the welfare loss, per car sold, for a 104 kg weight increase as: 104 * 0.0009 * (2.2/1,000) * $7 million = $1441 per car that changes weight in response to the policy. Our data suggest that 9% of cars increase weight and bunch at a notch point. The Japanese car market sells around 5 million new cars per year, so we estimate our aggregate annual welfare distortion to be 9% of 5,000,000 times $1441, which is $648 million.

If we use our lower bound estimate for the change in weight, the calculation will be changed as follows. We calculate the welfare loss, per car sold, for a 15 kg weight increase as: 15 * 0.0009 * (2.2/1,000) * $7 million = $208 per car that changes weight in response to the policy. Our data suggest that 9% of cars increase weight and bunch at a notch point. The Japanese car market sells around 5 million new cars per year, so we estimate our aggregate annual welfare distortion to be 9% of 5,000,000 times $208, which is $94 million.

4 Conclusion

This paper explores the economic implications of attribute-based regulation. We develop a theoretical framework that highlights conditions under which attribute basing is inefficient, and we show that, under those conditions, the use of attribute-based regulation leads to distortions that are concentrated in the provision of the attribute upon which targets are based. The model also explores cases where attribute-basing may play a role, but emphasizes that, even in those cases, the attribute-basing function deviates fundamentally from those observed in real world policies.

Empirically, the paper demonstrates that distortions in response to attribute-based fuel economy standards in Japan are clearly present. We use both established cross-sectional tools based on the notch literature, as well as novel panel techniques, to demonstrate that the Japanese car market has experienced a notable increase in weight in response to attribute-basing.

The theoretical framework makes a number of simplifying assumptions that could be relaxed in future work. In particular, it is possible that attribute-based regulation, particularly when it features notches, impacts firm pricing strategies for certain types of products. Further exploration of the cases in which attribute-basing may be justified is also warranted. On the empirical side, evidence of the responsiveness of attributes other than vehicle weight, which may be particularly easy to manipulate, would be a valuable object of study for future research.

References

Anderson, Michael and Maximilian Auffhammer. Forthcoming. "Pounds that Kill: The External Costs of Vehicle Weight." *Review of Economic Studies* .

Anderson, Soren T., Ian W.H. Parry, James M. Sallee, and Carolyn Fischer. 2011. "Automobile Fuel Economy Standards: Impacts, Efficiency and Alternatives." *Review of Environmental Economics and Policy* 5 (1):89–108.

Blinder, Alan S. and Harvey S. Rosen. 1985. "Notches." *American Economic Review* 75 (4):736–747.

Borenstein, Severin. 2013. "A Microeconomic Framework for Evaluating Energy Efficiency Rebound and Some Implications." Energy Institute at Haas Working Paper 242R.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *The Quarterly Journal of Economics* 126 (2):749–804.

Gillingham, Kenneth. 2013. "The Economics of Fuel Economy Standards versus Feebates." Manuscript: Yale University.

Holland, Stephen P., Christopher R. Knittel, and Jonathan E. Hughes. 2009. "Greenhouse Gas Reductions Under Low Carbon Fuel Standards?" *American Economic Journal: Economic Policy* 1 (1):106–146.

Ito, Koichiro. Forthcoming. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* .

Jacbosen, Mark R. 2013. "Fuel Economy and Safety: The Influences of Vehicle Class and Driver Behavior." *American Economic Journal: Applied Economics* 5 (3).

Jacobsen, Mark R. and Arthur A. van Benthem. 2013. "Vehicle Scrappage and Gasoline Policy." NBER Working Paper 19055.

Kleven, Henrik J. and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *The Quarterly Journal of Economics* 128 (2):669–723.

Knittel, Christopher R. 2011. "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector." *American Economic Review* 101 (7):3368–3399.

Metcalf, Gilbert E. 2009. "Tax Policies for Low-Carbon Technologies." *National Tax Journal* 62 (3):519–533.

Saez, E. 2003. "The effect of marginal tax rates on income: a panel study of bracket creep." *Journal of Public Economics* 87 (5-6):1231–1258.

Saez, Emmanuel. 1999. "Do Taxpayers Bunch at Kink Points?" *National Bureau of Economic Research Working Paper Series* No. 7366.

———. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3):180–212.

Saez, Emmanuel, Joel B. Slemrod, and Seth H. Giertz. 2012. "The elasticity of taxable income with respect to marginal tax rates: A critical review." *Journal of Economic Literature* 50 (1):3–50.

Sallee, James M. 2011. "The Taxation of Fuel Economy." *Tax Policy and The Economy* 25:1–37.

Sallee, James M. and Joel Slemrod. 2012. "Car Notches: Strategic Automaker Responses to Fuel Economy Policy." *Journal of Public Economics* 96 (11-12):981–999.

White, Michelle J. 2004. "The "Arms Race" on American Roads: The Effect of Sport Utility Vehicles and Pickup Trucks on Traffic Safety." *Journal of Law and Economics* 47 (2):333–355.

Whitefoot, Katie and Steven Skerlos. 2012. "Design Incentives to Increase Vehicle Size Created from the U.S. Footprint-based Fuel Economy Standards." *Energy Policy* 41:402–411.

A Our tax model is identical to a regulation

A majority of the attribute-based policies in the real world are regulations, but we wish to study them through the lens of optimal tax policy in order to facilitate comparison to Pigouvian taxes. We demonstrate here the equivalence between regulations and tax policies.

We first specify a fleet-wide energy efficiency regulation in the standard way. Consider a firm that produces goods, indexed by $j = 1, ..., J$ with energy consumption g_j, the sales-weighted average of which is required to exceed a standard σ. Firms can choose each good's price p_j, energy consumption g_j and some other attribute a_j. The firm solves a Bertrand competition problem over all of the products in its portfolio:

$$\begin{aligned} \max_{g_j, a_j, p_j} \pi = & \sum_{j \in J} (p_j - c_j) q_j \\ \text{s.t. } & \sum_{j \in J} \frac{q_j}{Q} g_j \leq \sigma, \end{aligned}$$

where q_j is the quantity sold, which is a function of g, a and p for all vehicles sold in the market; c_j is the marginal cost of vehicle j, which is a function of g_j and a_j but assumed to be constant in q_j, and $Q \equiv \sum_{j \in J} q_j$ is the total sales of the producer across all models. The Lagrangean for this problem is written:

$$\max_{g_j, a_j, p_j} \mathcal{L} = \sum_{j \in J} (p_j - c_j) q_j + \lambda Q \left(\sigma - \sum_{j \in J} \frac{q_j}{Q} g_j \right) = \sum_{j \in J} (p_j - c_j) q_j - \lambda \sum_{j \in J} q_j g_j + \lambda Q \sigma \tag{15}$$

where we multiply the constraint by Q so λ is interpreted as the per-unit shadow cost of tightening the regulatory constraint. This is the standard representation of a fleet efficiency standard.

Alternatively, suppose that there was a tax (or subsidy) equal to $t(g - k)$ per unit sold. That is, there is a marginal tax on fuel consumption g, but we allow also that this marginal tax is shifted uniformly by k so that the net tax on a vehicle could be positive or negative. The profit maximization problem for the firm in this case is:

$$\max_{g_j, a_j, p_j} \pi = \sum_{j \in J} (p_j - c_j) q_j - \sum_{j \in J} q_j (t g_j - k) = \sum_{j \in J} (p_j - c_j) q_j - t \sum_{j \in J} q_j g_j + tQk. \tag{16}$$

If $t = \lambda$ and $k = \sigma$, so that the tax on a product j is $\lambda(g_j - \sigma)$, then the maximization problem for the regulation (equation 15) and the tax problem (equation 16) are identical. Thus, the non-attribute-based regulation is identical to a tax on g.

In an attribute-based regulation, the target efficiency level of each vehicle is a function of the attribute, $\sigma(a)$. The standard that the firm must meet is the average target of its products, which is equivalent to saying that the firm's products must exceed their targets, on average. The Langragean under attribute-based regulation can be written as:

$$\begin{aligned}\max_{g_j,a_j,p_j} \mathcal{L} &= \sum_{j\in J}(p_j - c_j)q_j + \lambda Q\left(-\sum_{j\in J}\frac{q_j}{Q}\left(g_j - \sigma(a)\right)\right) \\ &= \sum_{j\in J}(p_j - c_j)q_j - \lambda\left(\sum_{j\in J} q_j\left(g_j - \sigma(a)\right)\right). \end{aligned} \tag{17}$$

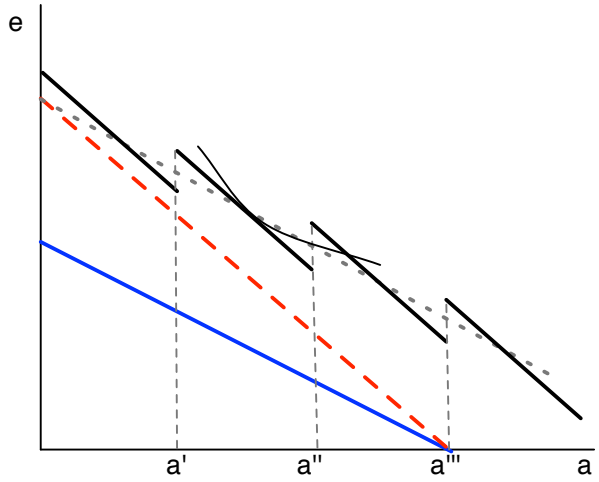
Alternatively, the attribute-based tax policy would be one in which each vehicle faced a tax rate of $t(g - \omega(a))$. In this case, the optimization problem for the firm is:

$$\max_{g_j,a_j,p_j} \pi = \sum_{j\in J}(p_j - c_j)q_j - t\sum_{j\in J} q_j(g_j - \omega(a)). \tag{18}$$

It is immediately apparent then that, when $t = \lambda$ and $\omega(a) = \sigma(a)$, the attributed-based regulation (equation 17) and tax (equation 18) problems are equivalent.

We specified here the problem faced by a single firm. With multiple firms, a common policy function $\sigma(a)$ will correspond to different shadow prices λ, and therefore a different equivalent tax policy. In fully tradable systems (such as CAFE), in which a firm that exceeds the standard can sell its excess credits to another firm, a single shadow price will prevail for all firms, provided that firms do not fail to trade for competitive reasons or because of illiquid markets or transaction costs. Thus, in our baseline model, the modeling of a tax that is uniform across firms embodies the assumption that there is full tradability and no competitive considerations prevent efficient exchange of credits across firms.

Figure 10: Isocost curve with notched attribute-based tax



B Notched policies have similar welfare properties

Many of the attribute-based regulations that we consider feature notches. That is, the function relating the fuel economy target to the attribute, $\sigma(a)$, is a step function with discrete jumps. Sallee and Slemrod (2012) study notched fuel economy policies and conclude that the use of notches in place of a smooth Pigouvian tax is welfare decreasing because it provides overly large incentives for fuel economy improvements for some agents and too small incentives for others. Our context is different in that we have two dimensions, and either could have notches.

We first consider the case where the notches exist only in the attribute function σ, but the tax on e is smooth. That is, conditional on e, the tax policy function $t(a, e)$ takes discrete jumps at certain values of a. How would such a policy function affect consumer choice? We can get some initial intuition graphically. Figure 10 shows an isocost curve, that is, the set of values for which a consumer spends a constant amount on the durable, $P(a, e) + t(a, e)$. The figure is drawn with several notches, at a', a'' and a'''. The solid blue line (drawn to be linear for the sake of illustration) shows the isocost curve before any policy intervention; and the dashed red line shows the modified isocost curve for the same expenditure on the good when there is a Pigouvian subsidy on e that has no attribute slope.

Next, the dashed grey line represents the isocost curve that would exist under a smooth attribute policy. In the diagram, the grey line is drawn parallel to the original blue line, which represents

the case when policy makers draw the attribute slope to match existing isocost curves, thereby preserving the original relative prices of a and e. This grey dashed line is not the final isocost curve, however, when $\sigma(a)$ is notched. In that case, the solid black lines represent the isocost curve for the consumer.

Importantly, the line segments on the final isocost curve are parallel to the red dashed line representing the Pigouvian tax. As in the smooth case, the existence of the attribute function does not distort the price of a relative to x, which means that the distortion in the choice of e will be only the indirect change due to a—it will be driven only by the utility and cost interactions of the optimal choice of e and the distorted choice of a. Furthermore, because the line segments are parallel in slope to the original Pigouvian line (and because we assume quasi-linearity) the choice of a will not be distorted if the consumer is choosing an interior point along one of the line segments. All of the distortion is due to cases where a consumer chooses a'. That is, all of the distortion is evident from those who "bunch" at the notch points. This guides our empirical work below.

We now provide algebraic analysis to flesh out the graphical intuition. For notational ease, we focus on the case with only one notch, at a', above which the tax subsidy jumps by amount $\tau > 0$. Then, the tax function can be written as:

$$t(a,e) = \begin{cases} t \cdot e & \text{if } a < a' \\ t \cdot e + \tau & \text{if } a \geq a'. \end{cases} \tag{19}$$

As above, denote by (a^*, e^*) the bundle chosen by a consumer facing a Pigouvian tax of $t \cdot e$. If the consumer's choice under the smooth attribute policy had $a^* > a'$, then the addition of the notch τ is purely an income effect. It has not changed the marginal price of a or e relative to each other or relative to x. Given quasi-linearity, this means that the durable choice of a consumer with $a^* > a'$ is unaffected by the introduction of a notched attribute policy.

When $a^* < a'$, the consumer will face a discrete choice of maintaining their original allocation or switching to a' exactly. They will not choose $a > a'$. To see why, suppose that they chose a value under the notched policy, call it $\tilde{a}$ strictly greater than a'. Then their optimization problem can be written $\mathcal{L} = U(a,e) + [I + G - P(a,e) + t \cdot e + \tau] + \mu[a - a']$, where there is a budget constraint as well as an inequality constraint that $a \geq a'$. If $\tilde{a} > a'$, then the shadow price on the latter constraint,

μ, is zero. In that case, the first-order conditions of the problem will be exactly the same as in the benchmark case with no attribute notch, which by construction featured an optimal choice of $a^* < a'$.

Thus, the consumer with $a^* < a'$ will either choose $\tilde{a} = a^*$ (and not receive τ) or will choose $\tilde{a} = a'$ exactly. This has the empirical implication that all bunching should come "from the left"—changes in a in response to the notched incentives should always be *increases* in a.

If a consumer chooses a', then their choice of e will solve:

$$\max_e = U(a', e) + [I + G - P(a', e) - t \cdot e - \tau], \tag{20}$$

which has the first order condition $\frac{\partial U(a',e)}{\partial e} = \frac{\partial P(a',e)}{\partial e} + t$. This compares to $\frac{\partial U(a^*,e)}{\partial e} = \frac{\partial P(a^*,e)}{\partial e} + t$ for the Pigouvian case, where $a' > a*$. Thus, just as in the smooth case, the distortion in e depends only the relative second-order curvature of P and U, which will be working in offsetting directions. If they offset perfectly, then $\tilde{e} = e^*$, but this would owe to a coincidence regarding the shape of P and U. More generally, the consumer's choice of e could rise or fall relative to the first-best choice.

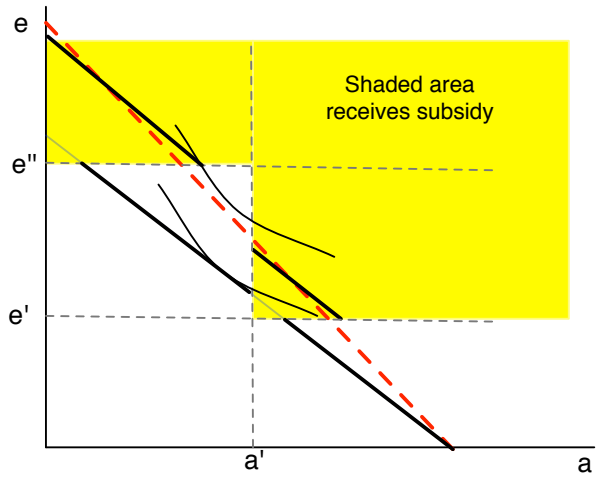
The distortion in a will be analogous to a traditional Harberger triangle and thus rising in τ^2. The consumer will choose $\tilde{a} = a'$ if and only if:

$$-\tau > P(a', \tilde{e}) - P(a^*, e^*) - (U(a', \tilde{e}) - U(a^*, e^*)), \tag{21}$$

that is, whenever the tax benefit is larger than the cost increase from moving from (a^*, e^*) to $(a', \tilde{e})$ minus the increase in utility from that change. The welfare loss can be written as a Taylor expansion, which again has the same intuition as a traditional Harberger triangle.

For our purposes, the main point of this analysis is that, even when the attribute function is notched, the focus of welfare analysis should be on how the policy distorts the choice of a relative to the Pigouvian baseline, and that we should expect the distortion to result in bunching at exactly the notch points in a. For empirical purposes, notched policies are useful in revealing the distortion because it is generally easier to detect bunching at specific notch points than shifts over time in an entire schedule. For that reason, the Japanese fuel economy standards are an especially fruitful context for study.

Figure 11: Isocost curve with notches for both a and e



All of this applies to situations where the (implicit or explicit) tax on e is smooth, but the attribute-basing function is notched. This describes fuel economy regulations in Japan. It is also possible, however, that both the tax on e and the attribute function σ have notches. That is, $t(a, e)$ is discontinuous in both a and e. This case describes the Japanese tax subsidy programs.

The simplest version of this policy is one with a single cutoff for a, call it a' and a pair of cutoffs for e, call them e' and e''. The tax for such a system can be described algebraically as:

$$t(a,e) = \begin{cases} \tau_1 & \text{if } e > e'' \\ \tau_2 & \text{if } e'' > e > e' \text{ and } a > a' \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

An isocost curve for this case is shown in Figure 11. The unsubsidized budget constraint is drawn as a faint line. The final budget constraint is represented by the bold black line segments, which overlap in parts with the unsubsidized line. Allocations in the yellow shaded area receive some subsidy. The subsidy is equal to τ_1 for any allocation above e''. Note that there are large regions of dominance in this diagram, where a subsidized point that has more of a and more of e has the same cost to the consumer as an unsubsidized bundle.

In the diagram, the red dashed line represents the simple Pigouvian tax. The values of τ_1 and τ_2 are chosen in this case to match the average Pigouvian subsidy for the relevant line segments,

but this need not be the case. Note that, if it is the case, then $\tau_1 \neq \tau_2$. In many policy examples, $\tau_1 = \tau_2$, which may be suboptimal.

When there are notches in both dimensions, there can be bunching in the distribution of e, at e' and e''. Above we argued that any change in a caused by attribute-basing relative to the Pigouvian optimum would come from *increases* in a. But, in cases with notches in both dimensions, it is possible that responses to the policy will lower a by inducing bunching at e' or e''. This would occur for cases like those represented by the sample utility curve in Figure 11, where a consumer's response to the notched subsidy is to bunch at e''. In that example, the indifference curve that is tangent to the unsubsidized budget constraint features a higher initial choice of a than at the bunch point.

C Technology spillovers create a role for attribute-basing, but not the one given to it by policymakers

Attribute based standards eliminate (or at least limit) the ability of producers to improve energy efficiency by downsizing in a, instead requiring them to make improvements in the products by dialing down attributes other than a that are correlated with e, or by adding technology that improves e while holding a constant. This approach could be appealing if there are positive externalities related to technological advancement. If firms cannot fully appropriate the value of their technological improvements, then technological improvements may represent a spillover benefit to other firms. In this case, there will be too little technological innovation in the absence of policy, and attribute standards might be beneficial in spurring innovation (instead of downsizing).

A direct way to model technological spillovers is to allow some secondary externality to directly enter the welfare function. We develop here a very reduced form model of these spillovers by simply specifying that costs incurred over some reference level, meant to represent the current state of the art or the current production possibility frontier, induce a positive external benefit. We are developing a more complete model of this spillover process, but we strongly suspect that the basic insights of the simplistic model represented here will carry forward in a more rigorous framework.

Denote the frontier cost function (some reference cost level) as $C(\bar{a}, \bar{e}) \equiv \bar{C}$. When the firm

chooses a bundle inside of the frontier, that is $C(a,e) \leq C(\bar{a},\bar{e})$, then the firm creates no technology spillover. The idea in this case is that the good is being produced with already established techniques. If the firm must push costs beyond the current frontier to make a product with a "higher" bundle of a and e, this requires some additional investment that costs $\rho(C(a,e)-C(\bar{a},\bar{e}))$. We make the functional form assumption that this extra cost is linear and increasing in its argument, so that the extra cost can be written as $\rho(a,e) = \rho \cdot \max\{C(a,e) - C(\bar{a},\bar{e}), 0\}$ and $\rho > 0$. Conversely, there is some externality that benefits all of society from the technological advancement, which we write as $\gamma(a,e) = \gamma \cdot \max\{C(a,e) - C(\bar{a},\bar{e}), 0)\}$

The consumer's and planner's problems can be written to include this spillover. There will be first-order conditions for each of two cases, where $C(a,e) > C(\bar{a},\bar{e})$ and where $C(a,e) \leq C(\bar{a},\bar{e})$. In either case, the consumer will fail to recognize the externality γ and will recognize only the costs. The only difference in this problem from the baseline model is that there will be a kink in the cost function at $C(\bar{a},\bar{e})$ which will induce "bunching"—not at a particular level of a or e, but along a particular isocost curve of a and e.

When the first-best choice involves values of a and e below $\bar{C}$, then the consumer's and planner's problems will be identical to the baseline. It is only when the first best allocation lies above $\bar{C}$ that this problem yields different conditions, so we focus on that here. In this case, the planner's problem becomes:

$$\max_{a,e} W = U(a,e) + [I - C(a,e) - \rho \cdot (C(a,e) - C(\bar{a},\bar{e}))] + \phi \cdot e + \gamma \cdot (C(a,e) - C(\bar{a},\bar{e})).$$

The planner's first-order conditions of this problem are:

$$\begin{aligned}\frac{\partial W}{\partial a} &= \frac{\partial U}{\partial a} - (1+\rho-\gamma)\frac{\partial C}{\partial a} = 0 \\ \frac{\partial W}{\partial e} &= \frac{\partial U}{\partial e} - (1+\rho-\gamma)\frac{\partial C}{\partial a} + \phi = 0.\end{aligned}$$

The consumer's first-order conditions are:

$$\begin{aligned}\frac{\partial W}{\partial a} &= \frac{\partial U}{\partial a} - (1+\rho)\frac{\partial C}{\partial a} + t\sigma' = 0 \\ \frac{\partial W}{\partial e} &= \frac{\partial U}{\partial e} - (1+\rho)\frac{\partial C}{\partial e} - t = 0.\end{aligned}$$

To make the consumer's first-order condition for e equal to the planner's, the tax on e must be set at $t = -(\phi + \gamma\frac{\partial C}{\partial e})$. This again is directly consistent with the Pigouvian tax result, where now the total externality related to e includes both the energy externality ϕ and the spillover externality γ. If $t = -(\phi + \gamma\frac{\partial C}{\partial e})$, then to make the consumer's first-order condition for a equal to the planner's, the attribute slope must be:

$$\sigma' = -\frac{\gamma\frac{\partial C}{\partial a}}{\phi + \gamma\frac{\partial C}{\partial e}}. \tag{23}$$

This result indicates that attribute-basing does have a potential role in the first-best solution when their are technology spillovers. Specifically, at the first-best, the slope of the attribute function should be equal to the ratio of the externalities associated with a and the externalities associated with e.

As we argued in section **??**, policymakers appear to commonly choose attribute-based schedules that follow the production frontier. That is, they choose a slope of $\frac{\partial C}{\partial a}/\frac{\partial C}{\partial e}$. This is optimal only if there is no externality related to e, and instead there is only a technological spillover, in which case $\phi = 0$ and equation 23 collapses to be only the ratio of marginal costs of the two attributes. More generally, the larger is the technology spillover γ relative to the energy externality ϕ, the closer will be the attribute slope to the production ratio. In contrast, as the energy externality is greater relative to the technology spillover ($\phi >> \gamma$), the attribute slope will approach zero.

Importantly, this result does not imply that attribute-basing is essential. Instead, attribute-basing provides a second instrument in addition to the tax on e. With two instruments and two externalities (one for e and one for a) it is possible to achieve the first best. That the two instruments are related via attribute-basing is not at all essential. The first best could equally well be achieved by subsidizing e at rate $\phi + \gamma\frac{\partial C}{\partial e}$ and subsidizing a at rate $\gamma\frac{\partial C}{\partial a}$.