Self Regulation, Corrective Policy and Goodhart's Law: The Case of Carbon Emissions from Automobiles

Mathias Reynaert
Toulouse School of Economics, University of Toulouse Capitole

James M. Sallee
University of California, Berkeley
and the NBER*

July 10, 2016
Preliminary draft: comments welcome

Abstract

Goodhart's law suggests that when an economic measure becomes the target of regulation, its measurement accuracy is eroded by strategic manipulation. We model Goodhart's Law for externality-correcting policies. A firm sends a signal about some attribute, the true value of which determines consumer demand and causes an externality that can be mitigated through regulation. Absent regulation, gaming (false signals about the attribute) reduces consumer surplus by distorting choice and allowing firms to raise prices. An equilibrium without gaming can be sustained because in subsequent periods consumers shift demand away from firms that game. Corrective policy increases private costs. Gaming thus lowers cost, and consumers share in the resulting private surplus gain. This causes self regulation to break down, because consumers prefer firms that game. We then consider an empirical application: the regulation of carbon emissions from automobiles in the European Union. We document a sea change in the reliability of laboratory-based carbon emissions ratings that coincided with the roll out of regulations based on these test ratings. Using panel data on 27 million fuel station visits from tens of thousands of drivers, we estimate that the difference between on-road fuel consumption and official laboratory tests increased from 5% before the regulation to more than 40% by 2014. This implies that 75% of the improvement in fuel economy attributed to the policy interventions is in fact due to gaming. At a social cost of carbon of $40 per ton, the program generates $1.2 billion less in carbon mitigating benefits annually than a naïve estimate would suggest. Using an estimated demand model we confirm in simulations that consumers privately gain when firms game to evade regulation. We also find that firms have a much higher incentive to game, and follow other firms' gaming, when the corrective policy is in place.

Keywords: gaming, corrective taxation, environmental regulation, carbon emissions, automobiles, fuel economy
JEL: Q5, H2, L5

*We thank seminar participants at ENSAE-CREST, Fundação Getúlio Vargas, the Norwegian School of Economics, Resources for the Future, the Toulouse School of Economics, and UC Davis for helpful comments.

1 Introduction

Goodhart's Law is often stated as "when a measure becomes a target, it ceases to be a good measure."[1] This truism creates a regulator's dilemma that is particularly nuanced when the measurement plays a coordinating role for economic activity among private agents. In such cases, the existence of the regulation may reduce the efficiency of private transactions by altering the fidelity of the measurement itself. Thus, a regulation subject to gaming may not only fail to achieve regulatory goals, but may also lower private welfare from market transactions.

We take up this problem both theoretically and empirically, with a focus on the case in which regulation is intended to mitigate some externality. Our theoretical model considers a monopolist who sells a good to a representative consumer. The good has some attribute that both directly determines consumer demand and creates a negative externality. This attribute is not directly observable, however, so consumer demand and any regulation are based on a signal sent by the firm. We model costly lying (Kartik 2009); the signal is a combination of the true attribute and gaming, both of which are costly to the firm. Goodhart's Law for externality-correcting policies manifests when the introduction of corrective policy induces gaming.

We relate our theory to an empirical application: laboratory-based fuel consumption ratings for automobiles. Consumers are willing to pay more for vehicles with greater fuel economy; fuel consumption creates negative externalities through carbon emissions; and automakers can provide more fuel economy either by changing the true performance of their vehicle or by gaming the emissions test. Goodhart's Law for externality-correcting policies manifests when fuel consumption regulation causes automakers to game emissions tests.

In our model, absent policy, consumers dislike gaming. When the firm misrepresents its product, consumers mis-optimize, leading to a loss in consumer surplus that we call choice distortion. Gaming also allows firms to raise prices, which further reduces consumer surplus. Consumer incentives are thus misaligned with firm incentives regarding gaming. Assuming that gaming is eventually revealed after consumers have purchased and used a product, self-interested consumers can self regulate the market by lowering demand for products that have been gamed.

Corrective policy disrupts this self regulation. Regulation raises the cost of production. Gaming allows the firm to lower its costs, and this benefits consumers through lower prices. When this price effect dominates choice distortions from faulty information, consumers benefit from firm gaming, even when they are fooled by it. As a result, consumer demand will rise with gaming, so consumer demand no longer acts as a self regulating force, leading to Goodhart's Law. Collectively, consumers would prefer to prevent gaming when the regulation is set optimally, because the losses in private surplus would be outweighed by gains from mitigating the externality, but individually rational consumers will free-ride and demand the gamed products when they are available.

It is perhaps obvious that corrective policy aligns the incentives of consumers and firms to collude against regulation, though we are not aware of this point being made in prior literature.

[1] Goodhart's original concern was monetary policy (Goodhart 1981). A similar notion, also focused on monetary policy, is captured in the Lucas critique (Lucas 1976).

After all, the nature of corrective policy is to reduce private surplus in exchange for mitigating an externality. This implies that corrective policy will exacerbate enforcement and compliance problems in markets. Key to our model is the fact that the regulatory measure is of direct interest to consumers, which makes clear how self regulation is natural in the absence of policy intervention. Moreover, we demonstrate that this result depends on the instrument available to the regulator. For example, in the case of fuel consumption regulations, Goodhart's Law appears in response to fuel economy regulations and vehicle taxes that depend on emissions ratings, but it would not appear in response to a gasoline tax. Raising the price of fuel aligns the interests of the consumer with the regulator, thus preserving the market's capacity for self regulation (though it is possible that fuel stations could evade a fuel tax).

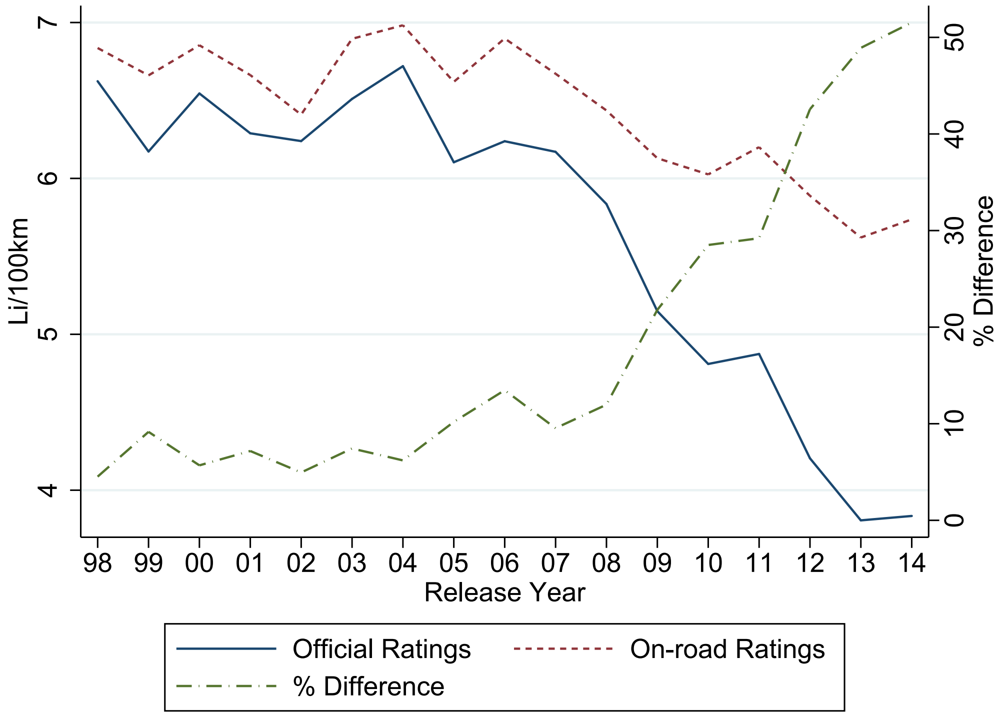
Our empirical analysis concerns the regulation of carbon emissions from automobiles in Europe. The E.U. recently passed a strict carbon emission standard for new vehicles that is similar to the Corporate Average Fuel Economy (CAFE) standard in the U.S. Both policies simultaneously regulate carbon emissions and fuel economy, where carbon emissions are assumed to be calculated as a fixed function of fuel economy ratings. The European law was initially passed in 2009, with a phase-in schedule that tightens the regulation over the next decade. In addition, a collection of nation-specific tax reforms changed between 2008-2010, which reformed registration and sales taxes so that they explicitly tax vehicles more heavily if they emit more carbon. All of these policies use measures of carbon emissions based on laboratory test. In recent years, media and industry reports have raised concerns over the accuracy of these laboratory tests in predicting on-road fuel economy. We interpret a gap between on-road performance and laboratory tests as a sign of strategic gaming.

We use a data set that records the fuel consumption and mileage of more than two-thousand types of cars driven by more than two-hundred and fifty thousand different drivers over twelve years in the Netherlands to measure the actual on-road fuel economy performance of new European cars, and we compare this to their official laboratory test ratings. We refer to the percentage difference between on-road performance and laboratory ratings as the *performance gap*. We find that vehicles produced before the introduction of the carbon regulation showed a small, and relatively stable, performance gap. Vehicles produced after the introduction of the standard exhibit a large and rising performance gap, so that 2014 model year vehicles now have performance gaps in excess of 45%.

A summary of our findings is presented in Figure 1, which plots the official rating, our estimated on-road performance, and the corresponding performance gap by release year (model year) of vehicles in our sample. The sea change in test reliability corresponds closely with the advent of the carbon regulation. We interpret this as evidence of strategic manipulation in response to the policy.

Our aggregate findings are consistent with industry analysis performed by the International Council for Clean Transportation (The International Council on Clean Transportation 2014 2015). They imply that around 75% of the apparent gains in fuel economy following the introduction of the regulation are eroded once the performance gap is taken into account. Extrapolating these

Figure 1: On-Road and official fuel consumption per release year



results to all of Europe and evaluating carbon emissions at $40 per ton, the difference between the apparent and actual emissions reductions amount to $1.2 billion annually from 2010 to 2014. In the empirical portion of our paper, we demonstrate that the findings are robust to a number of specification checks and controls. In all specifications we find a dramatic rise in the performance gap coincident with the introduction of the EU standard and a collection of nation-specific tax reforms, culminating in a fleet-wide average performance gap in excess of 45%. We find that the gap is not isolated to any particular automaker; a similar trend is apparent among all major automakers operating in Europe.

A growing set of media accounts detail the ways in which vehicle test ratings can be manipulated.[2] Tests are performed in third-party facilities, but these are funded by the automakers. The test procedures provides considerable latitude for automakers to modify the vehicle submitted for testing in ways that differ from the units sold to consumers. For example, test vehicles are fitted with special tires that have very low rolling resistance, which improves test performance. Automakers may also install particular technologies, such as stop-start engines, that are known to be more effective during test procedures than in real world driving. Finally, in the wake of the Volkswagen scandal, in which the firm admitted to using software to cheat on this same test for purposes of meeting nitrous oxide limits, it seems possible that automakers have used software to

[2]For example, see "Europe's Auto Makers Keep Test Firms Close" in the March 21, 2016 *Wall Street Journal*: http://on.wsj.com/1o5h47B.

cause a vehicle to perform especially well during the test.

The historical pattern in our data is consistent with our theoretical exposition of Goodhart's Law. The laboratory tests in question are the basis of both carbon regulations and consumer-facing fuel economy labels that are displayed by law on all new cars. As such, even before the standard was introduced, automakers could have benefited from gaming the test by improving their apparent fuel economy ratings to gain market share or increase prices. The absence (or moderation) of such gaming in the early 2000's, during which time fuel prices were historically high, suggests that some force, such as future losses from a damaged reputation among consumers, kept automakers from availing themselves of this opportunity. But, the implementation of a strict regulation and a suite of tax incentives moved automakers away from an honest equilibrium towards one with brazen gaming. Individual consumers (privately) benefited from gaming by avoiding costly price increases that would have been required by regulation.

To quantify the welfare consequences of gaming, we estimate a demand model of the European car market and use its structural parameters to simulate market outcomes from gaming with and without a policy. When there is no binding policy, and if consumers are completely unaware of gaming, we find that lowering perceived fuel costs by 5% of gaming reduces consumer welfare by €150 per car. This consumer loss goes directly to the gaming firms who are able to increase their profits. When we introduce policy we confirm that consumers gain from gaining. Despite similar choice distortions the total gain from avoiding a policy target that requires a 5% reduction in emissions amounts to €500 per car. We also find that the benefits to a sole firm gaming and the losses to a firm not gaming along with others triple compared to the no policy scenario. This shows that the pressure to game in the market clearly increases with the introduction of binding policy.

We contribute to several literatures. Our welfare analysis emphasizes the importance of multiple market failures, specifically how externality-correcting policies influence welfare through prices under imperfect competition. In doing so, it contributes to the literature started by Buchanan (1969), and recently advanced by Fowlie (2009) and Fowlie, Reguant, and Ryan (2016). Novel to this literature is our introduction of gaming and our analysis of how gaming interacts with market power to influence prices.

An existing literature in environmental economics does consider the challenges of compliance and enforcement (see Sigman (2012) for a review, and Sigman and Chang (2011) for a key example). Most recently, research has noted the importance of gaming or misrepresentation in automobile emissions programs (Oliva 2015) and in enforcement agencies (Duflo, Greenstone, and Ryan 2013). We differ importantly from that literature in emphasizing compliance issues among large multinational corporations operating in a wealthy country context. While some have stressed the potential importance of gaming and enforcement for choosing in environmental instrument choice (e.g., Goulder and Parry 2008), we provide a new analysis of how different instruments can influence compliance through incentives for market self regulation.

A long literature considers the economic efficiency of fuel-economy standards versus a gasoline tax (see Anderson and Sallee (Forthcoming) for a recent review). Recent articles have pointed out new concerns about the efficiency of regulation due to interactions with the used car market (Jacobsen and van Benthem 2015), safety (Jacobsen 2013), and heterogeneity in vehicle lifetimes (Jacobsen, Knittel, Sallee, and van Benthem 2016). We add to that literature by pointing out the importance of rating accuracy and gaming. This also contributes to a broader literature that has questioned the efficacy of energy efficiency policies more generally (e.g., Fowlie, Greenstone, and Wolfram 2015).

The balance of the paper is structured as follows. In section 2 we describe a theoretical framework that adapts Goodhart's Law to a setting in which private transactions are based on the target of regulation, and regulation is motivated by an externality. Section 3 describes the European automobile carbon emissions standard and the test procedure. Section 4 describes our data, and section 5 includes our main results. Section 6 calculates our estimates of lost consumer surplus, and section 7 concludes.

2 A model of gaming

This section develops a simple model to describe the determinants of strategic gaming and its welfare consequences. We begin with the simplest case, in which a monopolist sells a single product to a consumer in a single period. Many of our main insights can be obtained in this setting. We then discuss how introducing multiple time periods and oligopolistic competition affects interpretation. Our focus throughout is on understanding private surplus (excluding the externality), as our goal is to understand the incentives of individual firms and consumers to comply with a corrective policy.

2.1 Setup

We model a monopolist who sells a good that has an externality which depends on some mutable attribute of the good. The externality motivates regulation. The attribute that is regulated is of direct interest to the consumer and influences demand for the product. The attributes are not immediately verifiable, however, so consumers and the regulator must act on information provided by the firm. Deviation between the true attribute and information provided about it constitutes gaming.

These aspects of the problem are quite general and may reflect goods ranging from cigarettes to higher education. We make some additional assumptions motivated by the application we have in mind, where the attribute is energy efficiency of a durable good, that simplify exposition. First, where the attribute is efficiency, consumers care about efficiency to the degree that it affects operating costs, and we can thus model demand as determined by the full cost, including up-front price and lifetime operating cost. Second, motivated by media discussion of how emissions tests are gamed, we assume that gaming incurs only fixed costs for the firm; it does not change the marginal cost of the good.

The full cost of owning and operating the good is denoted $f = p + \beta x$, where p is the up-front purchase price, x is the energy consumption rating of the product and β is a coefficient that translates energy consumption ratings into dollars (e.g., for a vehicle, β is the price of fuel per liter times the number of present discounted lifetime kilometers driven, while x is the liters per kilometer rating). Consumers and the regulator do not observe x directly, but instead receive a message m from firms (e.g., the fuel economy label rating). Gaming occurs when a firm sends a message m that differs from x, where gaming g is defined as the difference between the true attribute and the message ($g = x - m$).

For simplicity, we assume that all attributes of a good are fixed other than x and p; this can be interpreted as a short-run assumption. The firm chooses x, p and g. We assume that the costs of gaming and of producing the true attribute are separable. The marginal cost of production for a product is $c(x)$, which is decreasing (x is a bad) and convex ($c' < 0$, $c'' > 0$). Consistent with media discussions of gaming in the automobile market, we assume that gaming is all fixed cost, which we denote $h(g)$, which is increasing and convex ($h' > 0, h'' > 0$).

Under full information, consumer demand D for the product depends on its various attributes (held fixed and thus suppressed in notation) and lifetime cost, which depends on x and p, as well as the lifetime costs and attributes of competing products. We write this as $D(f) = D(p + \beta x)$. Consumers, however, observe m not x, so they use the observed signal to form beliefs, labeled $\tilde{x}$, where $\tilde{f} = p + \beta\tilde{x}$. We assume risk neutrality.

Consumer beliefs about $\tilde{x}$ are assumed to be a weighted average of the truth and the belief that would occur if the firm gamed as much this period as it did last period. That is, $\tilde{x}_t = \alpha x_t + (1 - \alpha)(x_t - g_t + g_{t-1})$. When $\alpha = 1$, consumers are able to immediately "see through" gaming and infer the true attribute level from any signal. When $\alpha = 0$, consumers are myopic with respect to gaming and deflate any signal by the amount of gaming revealed from the prior period. Thus, their belief will be accurate ($\tilde{x}_t = x_t$) whenever the level of gaming is constant, but they will be mistaken when the firm games more (or less) than in the past. While somewhat restrictive, this assumption about beliefs nests a variety of possibilities. Here, a period can be interpreted as the length of time over which gaming is revealed through consumer experience with the product. We will focus initially on the one-period model, in which case we assume that $g_0 = 0$.

2.2 Policy

Policy intervention is motivated by a negative externality associated with x, which we assume is linear and equal to ϕx. We consider two instruments for regulating the message m, which is observable to the regulator, rather than the truth, which is unobserved. The first is a regulatory standard, which requires that fuel consumption be below a specified level σ: $m = x - g \leq \sigma$. We will use λ to denote the shadow price of the regulation *per unit*; i.e., the constraint on the firm's profit function is written $\lambda D(\sigma - x + g)$.

Second, as an alternative, the regulator could raise the price of fuel consumption through a tax on fuel τ. We model this as a percentage increase in β, so that fuel costs with a tax are equal to $(1 + \tau)\beta$. Note that either policy will end up affecting a product's full cost. A firm facing a regulation may change x and/or change price p, both of which affect demand only by shifting f.

Likewise, a fuel tax affects consumer demand by raising f.

Finally, while the regulator cannot observe the true attribute x at any moment in time, it may be able to observe it in later periods, or it may be able to conduct an investigation. Thus, we posit a penalty function that depends on gaming $r(g)$.

2.3 The monopolist's problem

The monopolist's optimization problem under a regulatory standard is:

$$\max_{p,x,g} \pi = (p - c(x))D(p + \beta x - (1-\alpha)\beta g) - h(g) - r(g)$$
$$\text{s.t. } x - g \leq \sigma.$$

Note that we maintain the implicit assumption that $g \geq 0$. We write the Lagrangean with the shadow price on the constraint scaled so that it is equal to the shadow cost per unit sold:

$$\mathcal{L} = (p - c(x) + \lambda(\sigma - x + g))D(p + \beta x - (1-\alpha)\beta g) - h(g) - r(g).$$

One key thing to note is that, absent policy, the monopolist will invest in x in a manner consistent with private efficiency. That is, if there were no externality, then the monopolist and the planner would choose the same level of x. When $\lambda = 0$, the monopolist will choose $-c'(x) = \beta$.[3] This says that they will lower x until the marginal cost of doing so equals the marginal value of increased energy efficiency to the consumer. The monopolist makes this privately efficient investment because any improvement in energy efficiency that is cost effective allows them to raise prices.

When there is a binding policy, it drives a wedge between the private cost and private benefit of energy efficiency. The monopolist will choose x to solve $-c'(x) = \beta + \lambda$. This is the standard "internalization" of the externality suggested by the Pigouvian tradition. If there were perfect compliance and there was no distortion to quantities sold due to market power, then the first-best solution would be obtained by setting σ so that $\lambda = \phi$.

2.4 The effect of gaming without a policy

Gaming has two effects on consumer welfare. To the degree that consumers (falsely) perceive gaming as a reduction in lifetime operating costs, gaming causes them to shift out their demand for any given up-front price p. This has two effects. First, firms will raise prices. Second, consumers will choose the wrong amount of the good, given their misperception about the full cost f. Both channels lead to a loss in consumer surplus.

To quantify the price effect, note that because efficiency enters demand in the same way as up-front price, gaming shifts out demand and triggers an equilibrium price change that is identical to the effect of a subsidy to consumers. But, because the gain is illusory, the welfare benefits are

[3] The first-order condition for x is $0 = -c'(x)D + (p - c(x))\beta D'$. Substituting in the optimal markup from the first-order condition for price ($p - c(x) = -D/D'$) yields the result.

the same as the effect of *taxing* the *firm* by the same amount; this is intuitive as both scenarios lead to no change in the good but an increase in price.

Specifically, consider a marginal increase in g, holding fixed the true attribute x. Consumers perceive a reduction in operating costs that depends on the value of efficiency β and the degree to which consumers are fooled by gaming α; i.e., $\partial f/\partial g = -(1-\alpha)\beta$ (where the partial derivative indicates that p is held constant as well as x). The firm prices strategically given this perceived demand curve. Up-front price p will thus rise by $\rho(1-\alpha)\beta$ for a marginal increase in gaming, where ρ is the pass-through coefficient that describes how consumer price changes in response to a tax. Weyl and Fabinger (2013) show that $\rho = (1 + (\epsilon_D - 1)/\epsilon_S + 1/\epsilon_{ms})^{-1}$, where ϵ is the elasticity of demand D, supply S and the inverse of the marginal surplus curve ms.

The good that consumers receive is unchanged, so consumer surplus can be evaluated using the original demand curve and the new price. Weyl and Fabinger (2013) show that the marginal decrease in consumer surplus is from a tax increase is $-\rho D$. To get our result, we need to scale the change in g to turn it into a change in the "tax" wedge, so $dCS/dg = -\rho(1-\alpha)\beta D$. We call this the price effect.

A second welfare effect comes from what we call choice distortion; the consumer misperceives the true full cost f of the product due to gaming and thus purchases too much of the good given its true ownership cost. This misoptimization creates a deadweight loss triangle that is equivalent to the consumer surplus portion of a Harberger triangle. Its width is the difference in demand, at the final price, induced by the gaming: $D(p + \beta x - (1-\alpha)\beta g) - D(p + \beta x)$. Its height is the perceived gap in fuel cost induced by gaming: $(1-\alpha)\beta g$. Thus, the area of this loss triangle is:

$$\begin{aligned}
\text{Choice Distortion} &= 1/2 \times \underbrace{(1-\alpha)\beta g}_{height} \underbrace{[D(p + \beta x - (1-\alpha)\beta g) - D(p + \beta x)]}_{width} \\
&\approx 1/2 \times \underbrace{(1-\alpha)\beta g}_{height} \underbrace{[-D'(1-\alpha)\beta g]}_{width} \\
&= -1/2 \times (1-\alpha)^2 \beta^2 g^2 D',
\end{aligned}$$

where the approximation assumes that the demand curve is locally linear. This is the traditional assumption invoked for a Harberger triangle. Where the demand curve has a large second derivative (curvature) and the amount of gaming is significantly away from zero, there will be an additional term ignored in this approximation. Choice distortion has the property common to Harberger triangles that it is zero when $g = 0$ and is rising linearly in the distortion g. The choice distortion also, intuitively, is larger when the demand derivative is larger. The derivative of the choice distortion with respect to g is $(1-\alpha)^2\beta^2 D' g$.

Importantly, in the absence of policy, both the price effect and the choice distortion work to lower consumer welfare. The consumer experiences a price increase, which lowers welfare. Then, they choose the wrong amount of the good according to their degree of misperception and their price sensitivity. Combining these effects yields our first result, which is that in the absence of policy, gaming lowers consumer welfare.

Proposition 1. *In the absence of policy, consumer surplus falls when gaming is introduced. Specifically:*

$$\frac{dCS}{dg} \approx -\underbrace{\rho(1-\alpha)\beta D}_{\textit{price effect}} + \underbrace{(1-\alpha)^2\beta^2 D' g}_{\textit{choice distortion}} \leq 0.$$

When consumers are fully sophisticated ($\alpha = 1$), gaming causes no change in consumer surplus; both terms go to zero. Both affects are larger when the degree of misperception $(1-\alpha)$ is larger. The price effect scales with the level of demand. The choice distortion scales with the slope of demand. Choice distortion is a tax wedge. It will be zero at the initial undistorted point ($g = 0$). Thus, at low levels of gaming, the price effect will be the dominant factor.

This analysis has proceeded assuming that x is fixed. This will in fact be the case in equilibrium in our setting, due to separability of the costs of x and g. The firm will choose x to satisfy $-c'(x) = \beta$, regardless of the choice of g. As a result, the true attribute of the product will in fact be unchanged when gaming is introduced. Absent policy, consumer surplus losses due to gaming come not from a distortion in x, but from price effects and misoptimization.

The profit maximizing level of gaming solves the following equation (derived by rearranging first-order conditions):

$$(1-\alpha)\beta D(p + \beta x - (1-\alpha)\beta g) = h'(g) + r'(g).$$

The marginal revenue of gaming (shift up in demand, $(1-\alpha)\beta$, times quantity, D) is equal to marginal cost, which can take the form of direct costs of gaming $h'(g)$, or expected regulatory penalties $r'(g)$. A firm will be honest ($g = 0$) when consumers are fully sophisticated ($\alpha = 1$), or when the regulatory penalty is sufficiently strong. In all other cases, the firm will game up until marginal revenue equals marginal cost, and this will incur welfare losses for consumers, coming both through the price and choice distortion channels described above.

2.5 The effect of policy without gaming when policy binds

Consider the imposition of a binding standard σ when there is no gaming. The firm must choose $x = \sigma$ to satisfy the regulation. This x is below the x that the firm would choose absent regulation (e.g., the vehicle has a lower fuel consumption rate), labeled x^p for private optimum. Define the change in x as $\Delta x = x^p - \sigma < 0$. The drop in x affects both demand and supply. Demand will shift up because the product is now less expensive to operate. Marginal cost will also shift up. What happens to prices and welfare?

The regulation can be understood as a pair of taxes. The regulation effectively provides a subsidy to the consumer of amount $\approx -\beta D'(p^p + \beta x^p)\Delta x > 0$. It simultaneously imposes a tax on the firm of amount $\approx c'(x^p)\Delta x > 0$. Were the subsidy and the tax equal in magnitude, the equilibrium quantity would be unchanged, as would consumer and producer surplus. Prices would rise by exactly the shift in fuel costs ($dp/dx = \beta$), so that f remained unchanged. Thus, full cost and quantity remain the same; so consumer surplus is unchanged. For the firm, marginal cost rises, but price rises by the same amount. With unchanged quantity, producer surplus is the same. [4]

A binding regulation will create a net tax on production, however. In the absence of policy, the monopolist's optimal choice of x equates $c'(x^p) = -\beta$. So, at a value of x below the private optimum, marginal cost will exceed benefits. Thus, a binding standard that reduces x to σ will cause a rise in marginal cost that exceeds the improvement in fuel costs experienced by the consumer.

The impact on prices of a marginal tightening of the regulation can thus be understood in two steps. A marginal tightening acts like a consumer subsidy equal to β, and a producer tax equal to c'. Decompose c' into the portion equal to β, and the portion that exceeds it in magnitude: $c' = -\beta + (c' + \beta) \equiv -\beta + w$, where w denotes the wedge. The portion of the tax c' that exactly offsets the β benefit to consumers will shift prices by exactly that amount: so $dp/dx = \beta$, and quantity will be unaffected. But, the second piece functions like a tax, creating a wedge equal to $c' + \beta$. This wedge acts exactly like a standard tax: so $dp/dx = \rho\beta$.

The net effect of a tightening of the standard on price is thus: $dp/d\sigma = \beta + \rho(c' + \beta)$. That is, prices will rise by the full amount of the drop in fuel costs, and it will rise by the pass-through rate times the wedge between $c' + \beta$. Note that the shadow price λ is equal to this wedge: $\lambda = \beta + c'$. So, the change in consumer surplus from the introduction of a binding standard is $dCS/d\sigma = (-\rho D)d\lambda/d\sigma$. The change in producer surplus is $dPS/d\sigma = -Dd\lambda/d\sigma$. This allows us to characterize the change in price, consumer surplus and producer surplus from a binding standard. The burden of the standard will be shared between producers and consumers according to the pass-through rate ρ. This pass-through rate depends on the elasticity of supply and demand, as well as the elasticity of marginal cost.

2.6 Gaming improves private welfare of consumers

Now, consider the introduction of gaming. First, suppose that consumers are fully aware of gaming ($\alpha = 1$). Then, for a binding subsidy σ, a marginal increase in gaming effectively lowers the tax wedge w. It is then obvious that the producer and the consumer *both* share in the benefits of gaming. Gaming lowers the tax wedge, and the benefits of this gaming will be shared by the producer and the consumer. That is, when the consumer is fully aware of gaming, the consumer will benefit from gaming in just the same way that they would benefit from a reduction in a tax.

When $\alpha < 1$, the consumer may still gain, but benefits from the price effect must outweigh the choice distortion. With a binding standard, any increase in gaming will cause a corresponding decrease in x: $dg = -dx$. That is, the firm substitutes compliance with g for compliance with x. This makes the analysis here different from the effect of gaming when there is no policy.

An increase in g with a corresponding decrease in x holds reported fuel consumption costs, but

[4] Another way to see this is to note that, because statutory incidence is irrelevant for a monopoly (Weyl and Fabinger 2013), the incidence of these "two taxes" is the same if we move the subsidy to consumers over to be a subsidy on producers, netting out the tax. If the tax and subsidy are the same magnitude, then the net tax is zero, so the equilibrium is unaffected.

not *perceived* fuel costs, which falls by amount $\alpha\beta$. In terms of demand, this acts like a tax on consumers in determining p.

Offsetting that is an effective subsidy to producers, who experience a drop in marginal cost equal to $c'(\sigma)$. Combining these two effects, up-front price will change as if there were a net tax to firms of amount $c'(\sigma) + \alpha\beta < 0$ (i.e., there is a net subsidy). The sign is determined because a binding standard implies that $c'(\sigma) + \beta < 0$, and $\alpha < 1$. This net subsidy will lead to a decrease in equilibrium up-front price, so that $dp/dg = \rho(c'(\sigma) + \alpha\beta)$.

This tells us that the up-front price change will be a decrease. But, the the true fuel cost also changes, so the effect on the full cost of the product is ambiguous. Specifically, the full cost changes by amount:

$$df/dg = \rho(c'(\sigma) + \alpha\beta) + \beta.$$

The first term is negative, but whether it is greater in magnitude than β is ambiguous. It is more likely to be negative as pass through is large. And, for a sufficiently high marginal cost (for a sufficiently tight standard), c' will be arbitrarily large, ensuring that the whole expression is negative. That is, when the standard is tight enough, the price effect of gaming will lead to a drop in the product's full cost, which benefits consumers. This can be true even when the consumer is completely fooled by gaming ($\alpha = 0$).

Gaming does introduce choice distortion here, just as in the situation without gaming. The choice distortion triangle has the same formula in either case. It depends on the gap between the true fuel cost and the perceived fuel cost, as well as the elasticity of demand.

Proposition 2. *In the presence of a binding standard, the introduction of gaming affects consumer surplus as follows:*

$$\frac{dCS}{dg} \approx \underbrace{(-\rho(c' + \alpha\beta) - \beta)D}_{\textit{price effect}} + \underbrace{(1-\alpha)^2\beta^2 D' g}_{\textit{choice distortion}} .$$

The choice distortion is the same as before: whether there is a policy or not, an increase in gaming causes an increase in the choice distortion. The price effect, however, is changed significantly, including a change in sign. As described above, the price effect can be negative, and it will be negative for a sufficiently tight standard. In turn, a negative price effect will dominate the choice distortion when the initial wedge is sufficiently large, which we state in the following corollary.

Corollary 1. *For a sufficiently stringent policy, a marginal increase in gaming will increase consumer surplus. That is, there exists some σ^* such that $\partial CS/\partial g > 0$ whenever $\sigma > \sigma^*$.*

This result has the flavor of a second-best reasoning in the presence of pre-existing distortions. The wedge between marginal cost and marginal benefit of x implies that an action, even a wasteful or distortionary one, that reduces that wedge will have a first-order benefit from reducing the wedge. This may offset the distortion that action creates directly. Here, there are two initial distortions: market power and the policy wedge. Thus, whether or not this second-best intuition holds depends on the nature of the market power distortion (depends on ρ), as well as the policy wedge.

Lastly, we state another corollary related to the Harberger triangle logic of an initial distortion, which captures the same idea. Whenever the price effect aids consumer surplus, the first unit of gaming will induce an increase in consumer surplus because the choice distortion will be small.

Corollary 2. *For a binding policy that induces a negative price effect, a marginal increase in gaming starting at zero will increase consumer surplus. That is,* $\partial CS/\partial g > 0$ *at* $g = 0$ *whenever* $\lambda > 0$.

To recap, when consumers are fully aware of gaming, the introduction of gaming benefits consumers because it lowers the policy wedge. The benefits of mitigating the wedge between private value and private cost is shared between consumer and producer in just the same way that an output tax would be. Even when the consumer is fully unaware of gaming, it is possible that they benefit. The full price of the product may go up or down, but if the initial wedge (the shadow price of the standard) is sufficiently large, then the consumer will benefit from lower realized price, despite their ignorance. The consumer will purchase the wrong amount of the product, however, because they perceive an even lower price. This choice distortion must be sufficiently small for the fully unaware consumer to benefit. Where the consumer is partially aware, the likelihood that they benefit on net from gaming rises.

Importantly, this discussion of consumer surplus ignores the externality. If the policy is set optimally by a planner who assumes perfect compliance, then by definition the representative consumer will benefit more from the externality reduction than they lose in private surplus. But, where individual consumers act according to their individual self interest, they will prefer to avoid the regulation where possible. Note also that if the policy is set too tightly, then consumer surplus can rise with gaming even accounting for the externality benefit. This appears to be true true in our empirical context.

2.7 Results differ under a fuel tax instead of a regulation

When the planner uses the fuel price instrument instead of the regulation, the firm's choices change in a way that alters the logic of the above findings. In raising the price of fuel, a tax on fuel inputs aligns the incentives of the consumer with the planner.

The firm's optimal choice of x is characterized by $-c'(x) = \beta(1+\tau)$. Thus, regardless of the degree of gaming, the firm will raise x to a higher level so as to meet a higher marginal benefit of efficiency.

Now consider the impact of gaming, starting from a case with a fuel tax, so $-c'(x) = \beta(1+\tau)$ and prices are made according to demand with that level of x and no gaming. Then, this situation is exactly the same as the case above where there is no policy at all, but replaced with a different value of β. That is, the result above showed that consumers lose from the introduction of gaming, for an arbitrary value of β in the absence of a binding regulation. The problem with a fuel tax is exactly the same as that problem, but with a new value of β. Thus, Proposition 1 above implies that, when there is a fuel tax as a corrective policy, the introduction of gaming causes a loss of private consumer surplus, both through a price effect and through choice distortion. The following corollary states this result.

Corollary 3. *When there is a fuel tax and no binding regulation, a marginal increase in gaming lowers consumer surplus. That is, $\partial CS/\partial g > 0$ whenever $\tau > 0$ and $\lambda = 0$.*

2.8 Market self regulation: a two-period model

The above propositions show a critical result that relates to what we call Goodhart's Law for externality-correcting policies. In the absence of a corrective policy, firm gaming lowers consumer welfare. As such, we say that the incentives of firms and consumers are misaligned. This makes market self regulation relatively easy; consumers have an incentive to punish firms that game. In contrast, when there is a binding standard, the firm and consumer both (privately, that is, not considering the externality) may benefit from gaming. We thus say that the policy aligns the interests of the firm and consumer together against the regulator. In a social sense, the consumer would like to avoid gaming so as to enjoy the externality mitigation (provided the standard is set optimally), but each individual consumer has an incentive to conspire and game.

This has a further implication about the ability of a market to self regulate. In the absence of policy, gaming only hurts consumers, so consumers would lower demand for a product suspected of gaming, which provides a natural correction within the market against gaming. This self regulation is lost when there is a corrective policy, however. The extreme case is when consumers are not fooled at all by gaming ($\alpha = 1$). In this case, the market fully self regulates in the absence of policy, but as soon as policy is introduced, the consumer conspires against the regulator.

We demonstrate this intuition more formally by introducing a second-period in our model, which we describe using subscripts to denote time period. A period of time in this model should be thought of as the length of time over which gaming becomes apparent. That is, we assume that in the second period, gaming in the first period is fully revealed.

The two-period Lagrangean can be written as follows:

$$\begin{aligned}\mathcal{L} &= \pi_1 + \pi_2 \\ &= (p_1 - c(x_1))D_1(\tilde{f}_1) - h(g_1) - r(g_1) + \lambda D_1(\sigma - x_1 + g_2) \\ &+ (p_2 - c(x_2))D_2(\tilde{f}_2) - h(g_2) - r(g_2) + \lambda D_2(\sigma - x_1 + g_2),\end{aligned}$$

where we supposed that the policy is the same in both periods. Substituting in the definition of consumer beliefs yields:

$$\begin{aligned}\mathcal{L} &= \pi_1 + \pi_2 \\ &= (p_1 - c(x_1))D_1[p_1 + \beta x_1 + (1-\alpha)\beta(-g_1 + g_0)] - h(g_1) - r(g_1) + \lambda D_1(\sigma - x_1 + g_1) \\ &+ (p_2 - c(x_2))D_2[p_2 + \beta x_2 + (1-\alpha)\beta(-g_2 + g_1)] - h(g_2) - r(g_2) + \lambda D_2(\sigma - x_2 + g_2).\end{aligned}$$

The firm chooses x_1, x_2, g_1, g_2, p_1 and p_2. We assume gaming in the prior (zero) period is zero.

2.9 Self regulation in the absence of policy

When will firms game in period 1? The first-order condition for g_1 is:

$$\frac{\partial \mathcal{L}}{\partial g_1} = -(p_1 - c(x_1))(1-\alpha)\beta D_1' - h_1' - r_1' + \lambda + (p_2 - c(x_2))(1-\alpha)\beta D_2' = 0.$$

Gaming in period 1 has several effects. First, it increases profits from the first-period demand by increasing demand by the amount $(1-\alpha)\beta D_1'$. Gaming has direct costs h_1' and potential regulatory penalty cost r_1'. It has regulatory benefits from relaxing the standard, equal to the shadow price λ. Finally, it changes beliefs about the product next period, which *lowers* demand in period 2 by amount $(1-\alpha)\beta D_2'$.

The first-order condition for price in period t implies that the markup is $p_t - c(x_t) = -D_t/D_t'$. Substitution yields:

$$\begin{aligned} 0 &= (1-\alpha)\beta D_1 - h_1' - r_1' + \lambda - (1-\alpha)\beta D_2 \\ &= (1-\alpha)\beta(D_1 - D_2) + \lambda - h_1' - r_1'. \end{aligned}$$

When might the market be self-regulating? If there were no policies at all, then $r_1' = 0$ and $\lambda = 0$. Then, the optimization condition simplifies to:

$$(1-\alpha)\beta(D_1 - D_2) = h_1'.$$

If consumers are fully informed ($\alpha = 0$) then there is no benefit to gaming; consumers see through the gaming immediately and it yields no demand boost. Assuming that gaming incurs some direct cost, no matter how small, implies that firms will not game.

Suppose that consumers are not fully informed, so that $\alpha > 0$. Then, self regulation will still occur whenever the market in periods 1 and 2 are the same size. Any increase in demand today is offset by a decrease in demand tomorrow.[5]

2.10 Policy disrupts self regulation

The introduction of a binding regulation disrupts the consumer's incentive to self regulate. This shifts the compliance burden fully onto the regulator.

Suppose that there is a binding policy, then the condition describing optimal gaming is:

$$(1-\alpha)\beta(D_1 - D_2) + \lambda = h_1' + r_1'.$$

In general, λ now creates a wedge, so that firms will game until the marginal cost of doing so is equal to the shadow price per product of the regulation, even when there is no market share advantage.

[5] Our model does not include discounting. Discounting would imply that some gaming occurs in the first period, just for moving profits forward in time. Note that this might be a substantial effect if a period is a long time. Recall that a period is the time over which gaming is revealed.

This extends the intuition from Proposition 2. Consumers benefit from gaming. Thus, even when the consumers are fully aware of gaming, the firm will game in equilibrium as a way to evade the regulator's intent, while maximizing the (private) desirability of the product to individually rational consumers.

For example, when consumers are fully aware of gaming, equation **??** reduces to $\lambda = h_1' + r_1'$. The marginal cost of the regulation will be equated to the cost of gaming, and the marginal cost of a penalty if one exists. In this case, it is also simple to show that $\partial g/\partial\lambda > 0$; as the regulatory shadow price rises, equation **??** implies that g will rise. As the regulatory stakes increase, the firm will game more. The generalized version of this is our final Proposition, which states Goodhart's Law for externality-correcting policies.

Proposition 3. *Goodhart's Law for externality-correcting policies: a tighter standard induces greater gaming:*

$$\frac{dg}{d\lambda} > 0.$$

Note that Goodhart's Law is not necessarily bad for consumer welfare, as implied by Proposition 2. But, it will reduce externality mitigation. That is, gaming trades-off consumer welfare and firm profits for less mitigation, which simply undoes the effects of the policy. In the meantime, it induces socials costs to the degree that gaming requires real resources.

This result does not hold for a fuel tax. A higher fuel tax does not induce more gaming.

Corollary 4. *Goodhart's Law does not apply to a fuel tax:*

$$\frac{dg}{d\tau} = 0.$$

Note that the clarity of this final corollary depends on the assumptions of our setup, in particular that the cost of gaming is independent of the cost of providing the true attribute. But, the qualitative result that a fuel tax is less likely to induce gaming than a regulation is likely quite general due to the difference in incentive alignments between the consumer, regulator and firm discussed throughout.

2.11 Introducing competition

Our empirical setting is characterized by imperfect Bertrand competition among differentiated products. For expositional clarity, the above treatment considers the case of a monopolist, but the results in Weyl and Fabinger (2013) indicate that the core results about price effects will hold for more general forms of competition, including our empirical setting. What changes is that the pass through depends on a conduct parameter that summarizes each firm's degree of market power. Thus, the formulas would change somewhat, but the basic intuition about price effects is generalizable.

Regarding choice distortion, the qualitative results from our monopoly case will also generalize. But, with multiple products, the choice distortion comes from switches between products as well as a reduction in the overall size of the market; that is, from the outside good. Moreover, some additional insights are possible in the case where there are multiple competing firms.

First, choice distortions will be larger when a single firm (or a subset of firms) game than when all firms game, unless the elasticity to the outside good is large. The reason is that, when all firms game by a similar amount, the relative price of products will change little. Thus, we might expect choice distortions to be small. The exception is if the outside good is very elastic, then a change in the overall price level brought about by gaming would cause consumers to purchase too much in aggregate.

Second, from the point of view of a single firm, gaming confers a competitive advantage. Conversely, when other firms are gaming, an honest firm would be at a competitive disadvantage. Thus, we expect competitive pressures to induce gaming. This is much the same intuition as in the monopolist; in either case the residual demand curve faced by the firm can be shifted out by gaming.

Below, we use estimates of the demand and marginal cost from a structural model in a simulation to calculate welfare impacts. The effects in our simulations relate directly to the components of welfare analysis that we describe above for the monopolist, but they are calculated under an assumption of Nash-Bertrand competition among firms with multiple differentiated products.

3 Fuel-economy regulation and the potential for gaming

Our empirical application concerns fuel-economy regulation in Europe. Like the Corporate Average Fuel Economy program in the U.S., the fuel-economy regulation in the E.U. requires automakers to sell vehicles each year that, sales-weighted, exceed a specified average fuel-economy level. Note that a fuel-economy standard, a fuel-consumption (the inverse of fuel-economy) standard, and a carbon-emissions standard are all synonymous. We use these terms interchangeably. The European policy is often described as a carbon-emissions standard, but the parallel program in the U.S. has always been referred to as a fuel-economy standard (though it now explicitly regulates both fuel economy and greenhouse gas emissions).

Firms can comply by improving the fuel-economy of their vehicles, or by shifting their sales mix to increase sales of models that are above the standard. Monetary standards for non-compliance are stiff; firms must pay 5 to 95 euros per unit out of compliance. When the standard was announced car-makers were facing fines of 1250 euros per vehicle if they would have not responded to the regulation.

The standard is weight-based; vehicles that are heavier are allowed to consume more fuel. This weight-basing limits the ability of automakers to comply through mix shifting by shrinking the variance in compliance status for individual models. This can be expected to push compliance actions towards technological improvements (Anderson and Sallee Forthcoming; Ito and Sallee 2015). Reynaert (2015) studies the early years of the program and finds exactly this; automakers have responded to the standard primarily through the adoption of improvements in the official fuel-consumption rating of vehicles, rather than greatly changing the composition of their fleet.

Prior to 2007, there was no legally binding fuel-economy regulation in Europe. High fuel prices have historically ensured a relatively efficient vehicle fleet in Europe. As part of the E.U.'s climate plan under the Kyoto Protocol, the E.U. established voluntary targets for fuel-economy improvements in conjunction with automakers. In the early 2000's, actual fuel economy improvements lagged behind the targets, and the E.U. replaced the voluntary program with a legally binding standard.

The standard was passed into law in 2009, and it allowed several years of phase in. The first year of enforcement was 2012, with a ramp-up in the standard taking place from 2012 to 2015. Fully phased-in, the regulation is quite aggressive by historical and international standards. On top of the EU standard all member states have separate taxation schemes for new vehicles. During the roll out of the EU emission standard many member states decided to adjust national policy in line with the EU target and focused taxation on emissions. France introduced a bonus-malus system, in place since 2008, that taxes polluting cars and subsidizes fuel efficient cars. Germany switched annual road taxes to depend on CO_2 rather than engine cylinders in 2009. Spain introduced registration taxes based on emission levels in 2008. The Netherlands did the same in 2010.[6] To measure fuel-consumption rates and emissions used to calculate compliance with the standard and national policies, the E.U. policy uses a driving test, which we describe in detail next.

3.1 The New European Driving Cycle fuel-consumption test

Fuel-consumption ratings in Europe are established by laboratory tests performed in third-party facilities that are funded by automakers. The test procedure, which is called the New European Driving Cycle (NEDC) measures fuel consumption in liters per 100 kilometers (l/100km). A test vehicle is put onto a chassis dynamometer (a treadmill for cars), and a professional driver "drives" the car through a specified series of speeds and accelerations. Emissions are captured directly from the tailpipe and used to determine gaseous outputs, which are used to determine fuel consumption. Two different test cycles are run to simulate different driving conditions. The first cycle resembles city traffic and consists of subsequent accelerations to low speeds (between 15km/h and 40km/h). The second cycle resembles urban traffic and consists of an additional, more gradual acceleration to 120km/h. Both cycles are fully standardized such that each car taking the test does exactly the same thing at each second of the test.

The NEDC is not only the basis of CO_2 regulation, but also the basis of consumer-facing information about fuel consumption ratings and the test used for determining emissions limits on local air pollutants, such as NO_X, PM and CO. The test procedure captures local pollutants

[6]See the ACEA tax reports (www.acea.be) for an overview of all vehicle taxation in the EU.

and measures their quantities to determine vehicle compliance with emissions limits. In terms of consumer information, the NEDC rating is the rating that automakers are required to use in consumer advertising, and it is the rating that appears on mandatory energy efficiency labels for new vehicles. As such, the NEDC has been used as a regulatory and market instrument for roughly a decade before the carbon standard was passed into law.

How might automakers game the test? According to media and industry accounts, the European test procedure offers the tester considerable "flexibility" in test procedures. For example, automakers are allowed to submit test vehicles that remove optional equipment, thereby changing the weight of the vehicle and improving performance. The test procedure involves a coast-down test to establish aerodynamic drag, during which a vehicle is put in neutral on a flat road at a specific speed. The regulation does not prohibit testers from taping down seams in the vehicle, removing side mirrors and roof racks, or over-inflating tires to improve rolling resistance during the coast down test. Alternatively, automakers may install technologies that perform particularly well on the test cycle, or they could even calibrate an engine to perform in a particularly efficient way during the highly specific test cycle's series of speeds and accelerations.

These "flexibilities" differ significantly from tests in the U.S. The U.S. fuel consumption tests are similar in spirit, though they use different driving courses. This may create a level difference in performance between the same vehicle tested under the two regimes. But, the difference in gaming likely stems not from the exact driving conditions required. Precise limits on the coast down test and the relationship between the test vehicle and production model are much more explicit in the U.S.

4 Measuring on-road carbon emissions

Our goal is to study the relationship between official test results and the true on-road performance of vehicles. Official test results are straightforward to obtain from industry data sources, but true on-road performance measurement is challenging. To do so, we obtained data for a large sample of drivers in the Netherlands that includes information on their fuel consumption and distance traveled. We use that to construct estimates of on-road performance, and then we estimate how that on-road performance changes with vehicle vintage.

4.1 Data

We obtained data from TravelCard NV, a company providing fuel services in the Netherlands. These panel data contain information on 66 million fuel station visits from drivers using a TravelCard NV card between January 2004 and May 2015. Most of the individuals in this sample drive a vehicle provided to them by their employer, who also pays directly for the fuel cost. This implies that we have a selected sample, though the provision of a company car is quite common in the Netherlands due to tax advantages and the high cost of personal vehicle ownership.

When visiting a fuel station, Travelcard NV users can swipe a smart card to pay for fuel. When a driver swipes her card we observe the drivers' license plate and the date, time and location of the fuel station visit. We also observe the exact amount and the type of fuel purchased and a self-reported odometer reading at the time of fueling.

TravelCard NV provides us with a second dataset that matches each license plate with the vehicle brand, model name, weight, fuel type and the official fuel consumption rating of the vehicle. These characteristics allow us to match the Dutch data with a panel on European car sales and prices from 1998-2011 used in Grigolon, Reynaert, and Verboven (2014) and Reynaert (2015). We match each vehicle to sales numbers, list prices and a broader set of characteristics containing length, width and additional engine characteristics.

The raw data include 66 million transactions, but many of the individual data points include unreliable self-reported odometer readings. Many odometer entries are missing, zero or apparently random entries. To deal with these data limitations, we take several steps to purge the data of unreliable observations, which cuts our final sample to 27.6 million transactions. In initial robustness tests, we find that censoring our data less severely increases our estimates of test gaming, which implies that, at least at first pass, our procedures are conservative against the finding of a test performance gap.

Specifically, first, we eliminate alternative fuel vehicles, or vehicles with missing fuel type information, and limit our sample to gasoline or diesel powered vehicles. This drops 7.1 million transactions. Second, we drop vehicles that use the wrong type of fuel for their engine in more than 1% of the visits, e.g., putting diesel fuel in a vehicle that is labeled as gasoline in our data. Inconsistencies might be in the data because drivers use their card for a different vehicle, or these observations might be mistakes in the assignment of vehicle type. This drops 7.5 million transactions. Third, we pose some minimum requirements on the driving patterns of the drivers that produce the transactions. We drop drivers that never report an increase of more than 150km in their odometer reading (2.5 million transactions).[7] We drop car models with less than 10 drivers, and drivers with less than 10 fuel station visits (1.3 million transactions). We drop drivers that did not report driving more than 5000km in total or reported driving more than 500,000km in total (11.3 million). Finally, we drop observations that result in a fuel consumption that is outside 1.25 times the interquartile range of estimated fuel consumption for each car model in the data (9.1 million). This results in the final dataset of 27.6 million observations.

The final data includes over 2500 unique types of cars driven by 275,000 different drivers. A car type is defined as a unique combination of brand (Volkswagen), model name (Golf), fuel-type (Diesel) and official fuel consumption. We define the release year for a car type as the first time we observe a unique combination of these variables in the data between 1998 and 2014.

Table 1 gives summary statistics for the raw data. The average vehicle in our sample has a fuel consumption of 6.65 li/100km and a weight of 1354kg. Somewhat less than half of the vehicles (45%) have diesel engines. We observe an average of 107 drivers per car with a maximum of 3227 drivers. For each of these drivers we observe an average 131 visits to the pump with a maximum

[7] Note that the range of a combustion engine is easily more than 800km.

Table 1: Summary statistics

	Mean	St. Dev.
Car Characteristics		
Fuel Consumption (Liters/100km)	6.65	1.73
Vehicle Weight in kg	1,354	245
Diesel Engines	0.45	0.50
Drivers per car	107	219
Driver Characteristics		
Pump visits	131	80
Total liters purchased	5,901	3,693
Total distance (km)	113,048	54,806
Pump Visit Characteristics		
Liters per visit	45.26	11.18
Odometer increase per visit	772	14,608

The table gives summary statistics for the 2696 vehicles, 275,640 drivers and 27,640,680 pump visits in the sample.

of 1138. Driver mean total consumption is 5,901 liters of fuel purchased corresponding to 113,048 km travelled. Finally, in the mean visit 45 liters of fuel is purchased and the odometer increases by 772 km. Notice that the standard deviation on the odometer is very large at 14,608. This is due to errors in the reporting of the odometer by drivers. Drivers need to type in the odometer on a pad in the fuel station and this is not always done with great precision. Below, we will discuss how we estimate on-road fuel consumption given this large variability in reporting.

An important question is how representative our sample is relative to the Dutch car market and by extension the whole EU. We find that the vehicles in our analysis cover 76% of sales on average between 1998-2011 the Netherlands. This means that, despite using a sample of company cars, we have information on most of the sales in the Netherlands. In Table 2 we compare characteristics of the vehicles in our sample with those of all cars in the Dutch market. We find that the mean vehicle in the Travelcard data is cheaper, lighter and more fuel efficient. Also, diesel cars have a higher share in our sample than in the market. One reason, for the lower means in characteristics is that our data contain almost no luxury and sports cars. Also, company cars are driven more than the average household car, which explains the higher share of diesels in our sample.

The Netherlands has a car market that is representative for the rest of Europe, with a balanced sales-mix of German, French and foreign brands. The NEDC driving test is valid across the EU and car makers do one test for the whole EU market. We want to stress that given the particular sample of company cars and their drivers it is unclear how representative the level of the estimated gap is. The average performance gap can be different for drivers without a company car. However, given the high sales coverage and robustness of our results, the changes in the performance gap over time can be seen as a EU wide phenomenon.

Table 2: Summary statistics: Netherlands and Travelcard

	Mean	St. Dev.	Mean	St. Dev.
	TravelCard		**Netherlands**	
Price (euro)	31,672	13,367	40,767	29,676
Fuel Consumption (Liters/100km)	6.74	1.60	7.89	2.46
Vehicle Weight in kg	1,344	230	1,409	308
Diesel Engines	0.45	0.50	0.36	0.48

Summary statistics for the TravelCard sample and the full dutch market between 1998 and 2011.

4.2 On-road fuel consumption

We construct a measure of on-road fuel consumption r_{nij} for each pump visit n of driver i in car j as the ratio of the liters purchased and the change in reported odometer between the visit and the previous visit:

$$r_{nij} = \frac{\text{liters}_n}{\text{odemeter}_n - \text{odemeter}_{n-1}} * 100 \tag{1}$$

This measure of on-road fuel consumption, in units of liters/100km, will vary between pump visits of a driver for three reasons. First, variable driving conditions such as outside temperature, route choice, driving style and congestion will differ across observations. Second, the driver may over or understate the odometer reading. We are not aware of any incentive for the drivers to deliberately misreport mileages but from the data it is obvious that there are many mistakes. Third, there might be variability due to stockpiling effects. If the consumer does not always fill the tank of the vehicle completely there will be variation in r_{nij}. If a driver visits the fuel station with an empty tank and fills half of it we will observe a very low fuel consumption for visit n and a higher fuel consumption for the next visit if she refills the tank completely.

Next, we construct the percentage gap between on-road and tested fuel consumption as:

$$d_{nij} = \frac{r_{nij} - li_j}{li_j} \tag{2}$$

in which the official rating li_j is constant for each car j and the on-road rating varies across observations.

We are interested in estimating the mean and variance of d_j, defined as the average d_{nij} across n and i for a given vehicle type j. In particular, we are interested in the mean and variance of d_j across vehicle types from the same vintage (release year). We take two approaches to estimating d_j. In a first exploratory approach we estimate regression equations with release year fixed effects, and we interpret those fixed effects as estimates of the average d_j across models from a given vintage, ranging from 1998 to 2014. In the regressions we will control for changes in driving behavior, sample selection and other factors that vary over time by including time (of driving) and vehicle type fixed effects. That is, we will observe a 2008 and a 2009 vehicle type both being driven in 2010.

In a second approach we follow the teacher value added literature to estimate on-road fuel consumption r_j using precision weights and an empirical Bayes correction. Our procedure follows Chetty, Friedman, and Rockoff (2014) and Kane and Staiger (2008), which provides a way for us to deal with substantial measurement error in odometer readings. To control for the large variance in reported odometer readings, and the variation in the number of drivers and visits observed for different types of cars, we construct a precision weighted mean of r_j and shrink it according to the reliability of the observations for j. Because of the measurement error in r_{nij} it is optimal, from a prediction standpoint, to use a biased but more precise estimate of each vehicle gap.

We start by attributing the total variance in the sample $Var(r_{nij}) = \sigma_r^2$ to three components: variance in performance of vehicles σ_j^2, drivers σ_i^2, and pump visits σ_n^2. We estimate the variance between fuel station visits of the same driver as:

$$\sigma_n^2 = \frac{1}{N-I}\sum_n^N (r_{nij} - \overline{r}_{ij})^2,$$

in which $\overline{r}_{ij}$ is the mean fuel consumption of driver i, N is the total number of observations and I is the total number of drivers. Next, we estimate the covariance between drivers of the same vehicles as:

$$\sigma_j^2 = cov(\overline{r}_{ij}, \overline{r}_{kj}),$$

in which we weight each pair of drivers (i, k) by the sum of their visits. We do this by randomly sorting drivers of the same car and then estimating the covariance between adjacent drivers. Finally, we obtain σ_i^2 as the remaining variance: $\sigma_i^2 = \sigma_r^2 - \sigma_n^2 - \sigma_j^2$. The precision of the estimated gap for each driver is then defined as:

$$h_i = 1/(\sigma_i^2 + \sigma_n^2/n_i),$$

so that drivers with a high number of visits have a higher precision. We obtain precision weighted means per car as the weighted average of $\overline{r}_{ij}$ with h_i as weights. Second we shrink these precision weighted means with an estimate of their reliability:

$$\psi_j = \sigma_j^2/(\sigma_j^2 + 1/\sum_i h_{ji}),$$

where the reliability is defined as the signal σ_j over the total variance. We use the per vehicle shrunken on-road estimates $\widehat{r}_j$ to construct an alternative estimate of the gap defined in (2).

5 Estimates of the degree of gaming

5.1 Mean fuel consumption and gap

In Figure 1 we plot the mean of official ratings $li_j, on-roadratingsr_{nij}$ and the %-gap d_{nij} for each release year. Between 1998 and 2008 we see that both official fuel consumption li_j and on-road consumption vary between 6.1 and 7 liters per 100km. The percentage difference between li_j and r_{nij} remains quite constant around 8%. From 2007 onward we see a spectacular drop in official consumption from 6.2 to less than 4 liters per 100km. This translates to a rise from the already high value of 38 mpg in 2004 to a truly remarkable 67 mpg in 2014. Official fuel consumption decreases by almost 50% over the sample period.

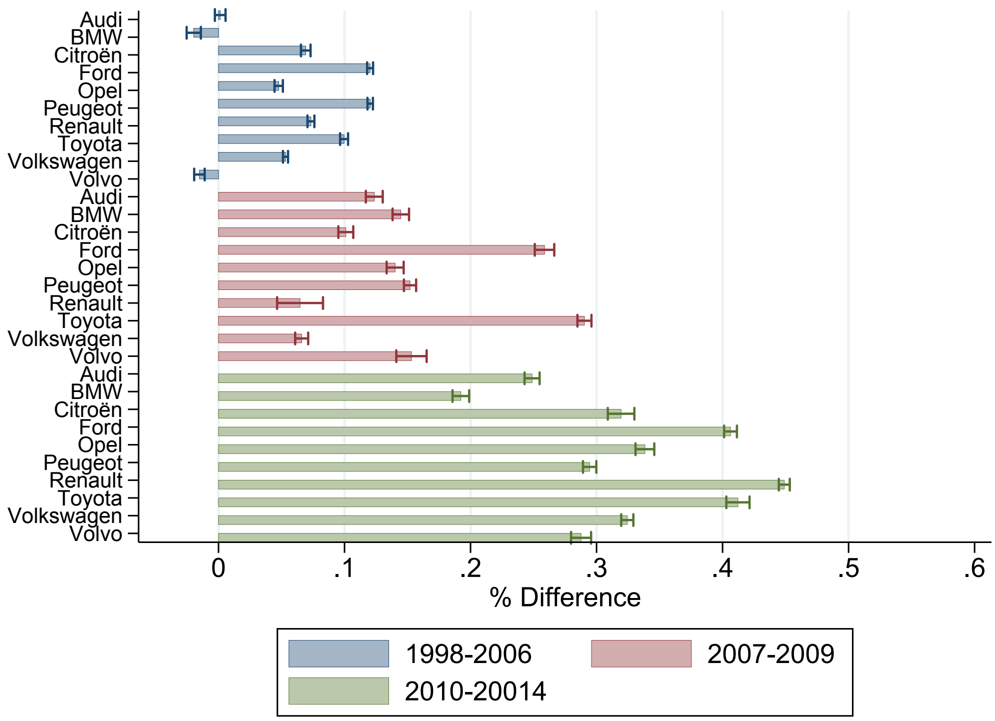
The on-road ratings follow a trend similar to the official ratings up until 2008. After 2008 the on-road fuel consumption decreases much more slowly than the official rating, going from 6.5 to 5.7 liters per 100km. Therefore, the gap between the official an on-road rating increases drastically after 2008 from around 10% at the beginning of the sample to almost 50% by 2014. This divergence is remarkable in magnitude, and it coincides exactly with the passing of the new fuel consumption regulations, which were passed into law in 2009 and phased-in over the remainder of the sample.

The divergence in test ratings and on-road performance is not isolated to a particular automaker. Figure 2 plots the estimated mean gap d_{nij} for vintage years until 2008 and vintage years after 2008 separately for each of the major automakers operating in Europe. All automakers show an average gap in excess of 20% over the final years of the sample, and all automakers exhibit a sharp increase in the performance gap over time. The gap is most pronounced for Ford, Renault and Toyota. Notice that these three firms have there roots in the US, EU and Japan, clearly showing that this is an industry wide issue not limited to European firms.

5.2 Fixed Effects Regression

Figure 3 plots the coefficients of two regressions with individual refueling transactions as the unit of observation. In the first regression (red markers) we explain the %-gap d_{nij} with release year dummies. We interpret the release year dummy variables as estimates of the mean performance gap among vehicles that were released in that year. The omitted category is vehicles that are present in the first year of our data, which implies that they were released in 1998 or earlier. In the second regression (blue markers) we also add the following controls to the explanatory variables: month and year of driving, fuel type and model type fixed effects. Full regression results are reported in columns (1) and (2) of Table 3. We add year and month dummies to control for variable driving conditions that affect all vehicles the same, given the time that they are driven. Congestion and routing have an important impact on on-road performance. If driving conditions are changing over time, not controlling for the conditions could bias our estimates because we observe later vintage year vehicles driven primarily later in our sample. We add fuel and vehicle model fixed effects to control for possible changes in the composition of the sample. If the gap is very different for different types of cars, than our vintage effects could represent a change in the fleet composition.

Figure 2: Gap between on-road and official fuel consumption per firm

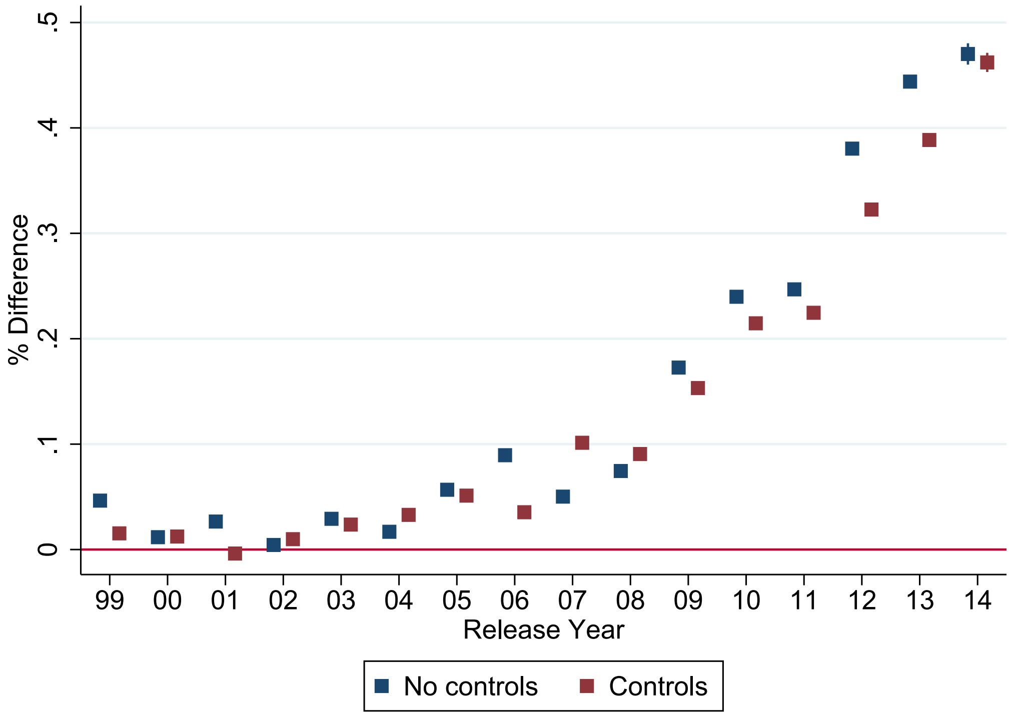


Company cars are subject to particular taxes and so over time the type of vehicles in the sample might change more swiftly than the overall car market.

The results show a gap that remains rather constant up to 2008 below %10 relative to 1998. After 2008 both release year effects from the regression with and without controls show that the gap increases to %45. Note that the difference between the estimated release year dummies with and without controls is small, such that the increase in the gap does not seem to be driven by driving conditions or sample composition. The R^2 for the empty model is 11%; given that the unit of observation is an individual refueling transaction, the release-year dummies explain an impressive fraction of the variation in d_{nij}. Our preferred interpretation is that this pattern across vintages implies a jump in strategic gaming in response to the roll-out of fuel consumption standards in 2009. What else might explain this pattern?

One possible alternative is a form of the rebound effect. If more recent vintages are more fuel efficient, consumers may respond by driving less carefully or using temperature controls or other equipment more often. In this case, reduced average fuel consumption rates will lead to an increasing gap. One might expect the same mechanism to create a significant difference in the gap between gasoline and diesel powered vehicles, as diesel vehicles are about 30% more energy efficient. Therefore, we estimate the release year fixed effects seperately for gasoline and diesel engines in columns (3) and (4) of Table 3. There is initially a 3% difference in the gap between diesels and gasolines, and this changes substantially over time. But, the coefficient patterns do not point to a

Figure 3: Release year coefficients from fixed effect regressions



larger gap among diesels commensurate with what would be implied by a rebound effect that could explain the magnitude of the pattern across vintages.

Finally, one might be concerned that the fuel consumption might differ with the age of the car. Typically, we expect older cars to become less fuel efficient leading to an overestimate of the earlier release year dummies. There might also be sample selection, however, such that cars with good on-road fuel economy stay long in our sample creating bias in the other direction. To control for this we only keep observations of d_{nij} that take place in the release year of the vehicle, so that we are capturing fuel consumption gaps only among the newest cars. Note that we are observing driving only from 2004 onwards, so that we can only estimate on-road performance from the year 2004 onwards. Column (5) reports those results, which show the same stark increase in the gap towards the end of the sample.

5.3 Variance decomposition and empirical Bayes estimates

In this subsection, we first decompose the variation in our data. We then provide an alternative description of the mean on-road performance gap that uses an empirical Bayes correction. Both procedures are described above in Section 4.2.

Table 4 describes the variation in the on-road fuel consumption (σ_r^2) across release years, as well as its decomposition across three components: variation across refueling transactions for the same driver (σ_n^2), variation across drivers of the same vehicle (σ_i^2), and variation across vehicles (σ_j^2).

Table 3: Release year fixed effects from regression of performance gap on controls

	All (1)	Controls (2)	Gasoline (3)	Diesel (4)	1st Year (5)
Constant	4.52	4.36	3.19	5.82	-0.68
	(0.09)	(0.12)	(0.13)	(0.12)	
1999	4.64	1.54	1.23	5.59	
	(0.14)	(0.13)	(0.27)	(0.15)	
2000	1.17	1.23	4.77	1.41	
	(0.15)	(0.15)	(0.25)	(0.18)	
2001	2.66	-0.38	0.93	3.38	
	(0.20)	(0.20)	(0.33)	(0.24)	
2002	0.43	0.99	-0.50	0.33	
	(0.21)	(0.19)	(0.36)	(0.25)	
2003	2.92	2.37	5.22	0.66	
	(0.21)	(0.19)	(0.29)	(0.30)	
2004	1.68	3.29	2.12	1.01	
	(0.15)	(0.16)	(0.24)	(0.19)	
2005	5.67	5.11	9.33	3.09	6.00
	(0.19)	(0.19)	(0.31)	(0.24)	(0.64)
2006	8.95	3.54	15.60	-1.24	11.7
	(0.28)	(0.32)	(0.36)	(0.42)	(0.55)
2007	5.02	10.1	6.47	3.63	6.78
	(0.20)	(0.26)	(0.30)	(0.26)	(0.39)
2008	7.45	9.06	6.46	8.13	18.6
	(0.24)	(0.25)	(0.34)	(0.34)	(0.57)
2009	17.3	15.3	17.2	18.3	21.3
	(0.20)	(0.22)	(0.28)	(0.28)	(0.49)
2010	24.0	21.5	16.5	27.5	25.7
	(0.17)	(0.20)	(0.28)	(0.20)	(0.56)
2011	24.7	22.5	23.5	25.0	29.2
	(0.18)	(0.23)	(0.33)	(0.21)	(0.43)
2012	38.0	32.3	29.8	44.2	43.9
	(0.19)	(0.24)	(0.26)	(0.23)	(0.37)
2013	44.4	38.9	40.0	45.2	51.6
	(0.26)	(0.26)	(0.38)	(0.33)	(0.35)
2014	47.0	46.2	46.7	49.8	49.4
	(0.52)	(0.46)	(0.36)	(1.62)	(0.58)
Year/Month F.E.		Yes			
Fueltype		Yes			
Model F.E.		Yes			
#Obs. (*10^6)	27.64	27.64	11.96	15.67	0.91
R^2	0.11	0.21	0.07	0.16	0.26

Standard errors clustered by driver in parentheses. Table reports coefficients from a regression of the performance gap (d_{nij}) on release year fixed effects. The unit of observation is an individual refueling transaction. Columns vary as follows: (1) contains all data, (2) all the data with year and month, fuel type and vehicle model fixed effects, (3) only gasoline engines, (4) only diesel engines, (5) keeps observations only in the first year of driving.

Table 4: Variance decomposition

	σ_r^2	σ_i^2	σ_j^2	σ_n^2
Mean	4.27	0.60	0.95	2.72
Standard deviation	1.50	0.24	0.32	0.99
Variance decomposition (%)	100	13.85	22.67	63.48
Standard deviation		1.30	4.82	5.09

σ_d^2 is the total variance in r_{nij}. σ_i^2 is the variation attributable to differences across individuals driving the same vehicle. σ_j^2 is the variation attributable to different vehicles. σ_n^2 is the variation across refueling visits of the same driver in the same vehicle. The variance decomposition is performed separately for each release year, and the mean and standard deviation across years are reported in table.

We decompose the variation separately for each release year and describe the mean and standard deviation across release years in the table.

More than 60% of the variance is attributable to within driver variance. This variance is due to driving conditions, stockpiling effects and errors in odometer reporting. We find that the variance across drivers of the same car σ_i^2 is 0.60. This is an economically large number; it means that the on-road fuel consumption is estimated to be 0.81 liter/100km higher at the third quartile than at the first quartile of drivers in the same car.[8] A policy that would shift a driver from the third quartile of the fuel consumption gap to the first quartile would decrease fuel consumption by 13%. These numbers are interesting from a policy perspective as they give an estimate of the extent to which fuel consumption and emissions can be reduced by teaching and incentivizing drivers to drive a vehicle more efficiently. There may be large gains attainable from changing driver behavior.[9] The remaining part of the variance σ_j^2 is the co-variance between drivers of the same car and can be seen as the information available to estimate the car specific component of on-road fuel consumption. We estimate this to be 0.95 or above 20 of the total variance.

Table 4 also shows that the variance components are relatively stable over time; each component has a low standard deviation across release years. There is variation in the size of the fuel consumption gap between cars and between drivers, but this variation is stable over time. Given this variance decomposition we turn next to the estimates of the distribution of r_j and d_j for each release year.

Table 5 reports the unweighted and uncorrected estimate of r_j (the simple mean per vehicle j) and $\widehat{r}_j$, which employs the empirical Bayes correction. The mean value of both r_j and $\widehat{r}_j$ are decreasing over the release years. In all years the corrected means are lower than the raw means, because on average vehicles with high r_j have less precise underlying data, but overall shrinkage and

[8]If we assume that conditional on car j, r has a normal distribution, the interquartile distance is $1.349 * \sigma_i^2$.

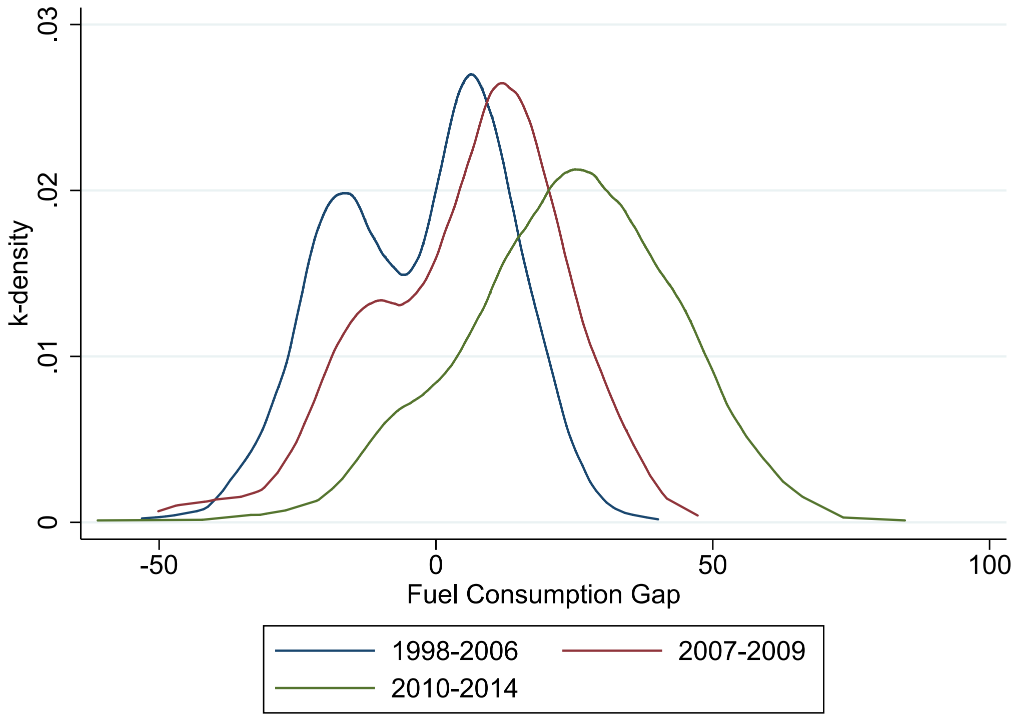
[9]Significant variation across drivers of identical cars is consistent with results reported in Langer and McRae (2014). Their data derive from a few dozen drivers of a single vehicle model, a Honda Accord, whereas our estimates come from tens of thousands of drivers observed over hundreds of different vehicle models.

Table 5: Estimated vehicle gaps per release years, with and without empirical Bayes correction

Release	Unweighted r_j	$\widehat{r}_j$	Shrinkage	$\widehat{d}_j$
1998	6.98	6.88	0.98	-4.82
	(1.23)	(1.18)	(0.02)	(14.71)
1999	7.11	7.02	0.98	-2.07
	(1.28)	(1.21)	(0.02)	(14.04)
2000	7.06	6.98	0.98	-3.86
	(1.33)	(1.31)	(0.02)	(14.37)
2001	6.99	6.88	0.98	0.11
	(1.20)	(1.17)	(0.02)	(13.46)
2002	6.70	6.55	0.97	-3.20
	(1.19)	(1.13)	(0.02)	(16.77)
2003	7.05	6.91	0.98	-0.78
	(1.38)	(1.34)	(0.02)	(15.94)
2004	7.24	7.11	0.98	-4.96
	(1.54)	(1.47)	(0.02)	(16.90)
2005	7.09	6.90	0.97	0.33
	(1.37)	(1.27)	(0.02)	(16.38)
2006	7.24	7.07	0.97	-0.76
	(1.35)	(1.29)	(0.02)	(15.59)
2007	6.77	6.61	0.97	2.52
	(1.14)	(1.08)	(0.02)	(17.26)
2008	6.67	6.53	0.96	4.13
	(1.18)	(1.14)	(0.03)	(16.56)
2009	6.54	6.40	0.98	9.62
	(1.16)	(1.13)	(0.02)	(16.30)
2010	6.52	6.41	0.99	14.77
	(1.18)	(1.15)	(0.01)	(17.64)
2011	6.43	6.33	0.98	19.75
	(1.15)	(1.14)	(0.01)	(16.48)
2012	6.18	6.10	0.99	27.11
	(1.09)	(1.00)	(0.01)	(16.61)
2013	6.09	6.03	0.98	40.02
	(0.91)	(0.87)	(0.02)	(12.89)
2014	6.00	5.97	0.95	49.99
	(0.89)	(0.84)	(0.05)	(16.32)

Standard Deviations in parentheses.

Figure 4: k-Density of vehicle fuel consumption gap



precision weighting has no significant effect. The resulting gap $\widehat{d_j}$ is estimated to be an imprecise zero up until 2008. From 2009 onwards we see a significant increase in the performance gap, consistent with the previous estimates. Note that the mean of the estimated gaps increases over time but its standard deviation remains relatively constant between 14% and %20. To see this, Figure 4 gives the kernel density of the estimated gaps for release years beween 1998-2006, 2007-2009 and after 2009. The densities show clearly that the standard deviation is relatively stable over time with the whole distribution shifting to the right because of the mean increase in the gap.

We conclude from this empirical section that there is very strong evidence for substantial gaming on fuel consumption ratings. Depending on our estimation method the gap between on-road and tested fuel consumption varies between 0% to 10% before regulation. By 2012 this gap has increased to 30% and all estimates give a gap of more than 40% in 2014. We do not find a single robustness check where this spectacular increase in the gap does not occur. Strikingly and somewhat against our expectations we find the gap to be increasing for all vehicles and brands in the sample. The distribution of the gap is shifting, but the shape is relatively constant. We now consider the implications of gaming for the evaluation of environmental policy.

6 Welfare analysis

We start by estimating a demand model for new vehicles in order to obtain taste parameters for the consumers. Given these estimated taste parameters, we simulate several counterfactual simulations that quantify changes in consumer surplus and profits with and without binding regulations on fuel consumption. We proceed by describing the estimation, simulation and computation of consumer surplus and end with the results of the counterfactual. Finally, we discuss the cost of missing the externality target of the regulation.

6.1 Estimation of the automobile demand system

Each consumer i chooses the vehicle that maximizes her indirect utility, which we write as:

$$u_{ij} = \Delta_j \gamma_i - \eta p_j + \beta_i x_j + \xi_j + \varepsilon_{ij},$$

where Δ_j is a vector of vehicle characteristics, p_j is price and x_j is the operating cost of the vehicle for driver i, measured as fuel consumption (in euro per kilometer). We estimate a random coefficient logit model such that we estimate the mean and standard deviations of β_i and γ_i under the assumption that consumer tastes are normally distributed. All remaining consumer heterogeneity is contained in the additive idiosyncratic error term. Following Berry, Levinsohn, and Pakes (1995), we integrate out ε_{ij} to construct choice probabilities. After inverting the choice probabilities to obtain ξ_j, we use a GMM estimator such that the vector of parameters θ is the solution of:

$$\min_{\theta} \xi_j Z' \omega Z \xi_j$$

in which Z is a matrix of instruments and ω is a weighting matrix. We use the data containing sales and prices from the Netherlands from 1998 to 2007. The vector Δ contains information on horsepower, weight, footprint (size of the base) and height. Additionally, we control for a linear time trend, brand dummies, body class dummies, number of doors and months of sales in the calendar year. We instrument for prices using both cost shifters and sums characteristics instruments as in Reynaert (2015). We set the share of the outside good equal to 5% so that most of the effects in the counterfactual are from changes among current buyers. We use data before 2007 to estimate the demand under the assumption that x_j represents true fuel costs and there is no variation in gaming before 2007 resulting in an unbiased estimate of β.

Estimated taste parameters and standard errors are reported in Table 6. Price and fuel costs have the expected negative effect on utility. We estimate considerable heterogeneity in both the taste for fuel costs and horsepower. We find results that are comparable to previous work using this data, see Grigolon, Reynaert, and Verboven (2014) and Reynaert (2015). Note that several parameters are estimated with a large standard error. The reason for this is that we only use data from the Netherlands and thus have a panel limited in size.[10]

[10] To increase precision, we plan to estimate the taste parameters using the larger EU sample with more markets

Table 6: Estimation Results

	Mean Taste		St. Dev.	
	Coeff.	St. Error	Coeff.	St. Error
Price	-5.51	(5.31)		
Fuel Cost	-0.42	(0.18)	-2.64	(3.45)
Horsepower	-1.56	(0.42)	4.16	(2.84)
Weight	4.02	(3.48)		
Footprint	1.13	(0.67)		
Height	-0.03	(0.15)		
Diesel	-1.48	(0.24)		
Doors	0.23	(0.11)		

6.2 Welfare estimates under alternative assumptions about policy and consumer sophistication

In this section we quantify the welfare effects of gaming when there is a binding policy ($\sigma > 0$) and when there is no binding policy ($\sigma = 0$). The simulations differ from the theory in the sense that we are solving for equilibria in a differentiated product oligopoly instead of a homogeneous product oligopoly. For each simulation we do the following. First, decrease the fuel cost of the gaming firm(s) by 5%. Second, solve for new demand and prices given the assumption that consumers are not sophisticated. For each scenario we also describe what happens when consumers are sophisticated. Third, compute changes in welfare relative to a base scenario without gaming.

Implicitly we make the following assumptions. First, gaming is a fixed cost: it does not shift the marginal costs of the firms. Second, firms will change their pricing when gaming if consumers are not sophisticated. Third, we simulate the policy as a 5% decrease that each firm has to obtain on its sales-weighted emissions.[11]. Fourth, we assume that consumers are do not see through gaming and thus base their choices on the stated fuel costs at the time of purchase. We simulate scenarios with 5% of gaming, roughly equilavent to the yearly change in gaming that we observe. In the theory this would be equivalent to $\alpha = 0$ and a feedback loop of 1 year. The obtained changes in welfare are the changes in yearly utility and profits from new vehicle sales. Next, we describe how we compute consumer and producer surplus.

6.2.1 Consumer Welfare and Profits

When consumers are sophisticated $\alpha = 1$ and the policy is not binding, gaming will not change consumer welfare. To compute consumer welfare when consumers are affected by gaming we follow Dubois, Griffith, and O'Connell (2016), who describe the welfare effects of persuasive advertising. As described in section ?? gaming increases prices from p_j^0 to p_j^1, as firms will exploit higher demand

to increase precision.

[11]We use the actual formula used by the EU in its emission standard such that the policy is attribute based on weight. For each firm we compute a sales weighted average emission, with emission for heavier (lighter) cars receiving a bonus (penalty) in the weigthed sum. See Reynaert (2015) for a detailed description.

from perceived lower fuel costs. Gaming will also distort choices as consumers wrongly perceive fuel costs. To separate the two effects we make a distinction between decision (at the moment of purchase) and experience utility (at the moment of utilization).[12]. When there is gaming, a naïve consumer will perceive fuel costs, following her belief, as $\widetilde{x}_j = d_j * x_j$, in which d_j is the difference in gaming between this and the previous period. The consumer will make her choice based on $\widetilde{x}_j$ and will perceive her decision as yielding a utility of:

$$\widetilde{V}_{ij}(d, p^1) = \Delta_j \gamma_i - \eta p_j^1 + \beta \widetilde{x}_j + \xi_j + \varepsilon_{ij}.$$

After purchasing the vehicle, true fuel costs are revealed, and the consumer has experience utility:

$$V_{ij}(d, p^1) = \Delta_j \gamma_i - \eta p_j^1 + \beta x_j + \xi_j + \varepsilon_{ij}.$$

The difference between decision and experience utility is the optimization error, which in this case is the value of the additional fuel costs for the consumer $\beta(x_j - \widetilde{x}_j) = \beta(1 - d_j)x_j$. Consumer surplus with gaming can then be written as:

$$\begin{aligned} \widetilde{W}_i(d, p^1) &= \mathbb{E}_\varepsilon[\widetilde{V}_{ij}] - \mathbb{E}_\varepsilon[\beta(1 - d_j)x_j] \\ &= W_i(d, p^1) - \sum_j [s_{ij}\beta(1 - d_j)x_j] \end{aligned}$$

where s_{ij} are the choice probabilities obtained from maximizing the decision utility. We compute $W_i(d, p^1)$ by applying the log-sum formula of Small and Rosen (1981). We can then decompose the change from an honnest equilibrium with $(0, p^0)$ to an equilibrium with gaming (d, p^1) into a price effect and choice distortion:

$$W_i(0, p^0) - \widetilde{W}_i(d, p^1) = \underbrace{W_i(0, p^1) - \widetilde{W}_i(d, p^1)}_{\text{Choice Distortion}} + \underbrace{W_i(0, p^0) - W_i(0, p^1)}_{\text{Price Effect}}$$

When there is a policy we will have exactly the same decomposition but the price effect will now consist of two parts: the shift in prices due to gaming and the shift in prices from lowering the tax wedge (a decrease in λ).

Profits of firms are given by:

$$\pi_f = \sum_j [(p_j - c_j)s_j(d, p)A]$$

in which A is the size of the market. Profits will depend on the margin and the demand, both a function of the price schedule chosen. The policy will force the firm to change the price schedule, while gaming will shift the demand and will also change the optimal prices of firms.

[12] Note that this is a similar framework as in Allcott (2013) and Sallee (2014) who study misperception of fuel costs but more general as we allow for firms to change prices when they game

6.2.2 Effects of gaming without policy ($\sigma = 0$)

Results describing the effect of gaming without a policy are given in the first panels of Table 7 and Table 8. We compute the changes in private welfare when all firms game, when one firm games or when all but one firm games. We decompose consumer welfare changes in a price effect and a choice distortion. When all firms game consumers loose €148 per vehicle. The price distortion is responsible for almost all of the effect, consumers pay more for products with lower perceived fuel costs. The choice distortion is small at only €6.56, if everyone games consumers are not very distorted as the relative ordering of vehicles remains more or less equal. Interestingly, when only one firm games the choice distortion is larger (€7.8) and the price effect a lot smaller. This makes sense, only one firm will be able to charge a higher price for lower fuel costs and by rising prices in more limited amounts the firm will steal market share. If all firms but one game the choice distortion is largest €12.67. The price effect is very similar as when every firm games.

In terms of profit changes we see very moderate effects if everyone games. The overall effect is that profits increase slightly but this is not true for BMW and Volkswagen. Because of the choice distortion consumers choose different vehicles at different fuel cost levels and both firms loose a bit of market share due to this reshuffling of shares. The profit gains from being the only gamer are very large: each firm wins between 11% to 17% from gaming. Contrary, being the only firm that does not game hurts profits by 10%-14%.

6.2.3 Effects of gaming with policy ($\sigma > 0$)

With a policy there is one additional effect of gaming: it reduces the regulatory constraint. We see this effect for consumers very clearly in the second panel of Table 7. The price effect changes sign and increases threefold to €489 if every firm games. This shows that consumers prefer the price schedule offered under gaming to that offered with honest compliance to a policy. The size of the choice distortions are very similar to those without policy. This leads to the conclusion that consumers are better off with gaming when the policy is binding. Note that if consumers would be able to perfectly see through gaming ($\alpha = 1$) we would see even more positive welfare effects as consumers only gain from policy avoidance and are not affected by increasing price effects or choice distortions.

The effects from gaming on profits are higher and more diverse with a policy than without a policy. The regulation does not have a negative effect on all firms, and so avoiding the regulation by gaming decreases their profits relative to complying honestly to the regulation. This is the case for BMW, Daimler and Ford.[13] All other firms gain from avoiding the regulation, Fiat is the strongest beneficiary with gains up to 13%.

Interestingly, the gains from being the only gamer and the costs of being the only honest firm are a lot larger with the policy in place. Being the only gamer increases profits by a more than 20%.

[13] Note that this is partly due to our estimation with a very low share of outside good. This results in a very low substitution towards the outside good when the regulation is imposed, keeping total sales more or less constant. If we would re-estimate the model with a higher outside good share, all firms would lose because of decreases in sales.

Table 7: Changes in Consumer Welfare from Gaming (Euros per vehicle)

	All Game	Own Gaming	All other game
		No Policy	
Total CS	-148.43	-18.60	-139.32
Choice Dist.	-6.56	-7.80	-12.67
Price Effect	-141.88	-10.80	-126.66
		Policy	
Total CS	489.88	40.48	443.17
Choice Dist.	-6.47	-8.23	-11.86
Price Effect	496.36	48.70	455.04

Column (1) gives changes in CS when all firms game, column (2) gives the average of the changes in CS when each firm games on its own, column (3) gives the average changes in CS when all but one firm games.

Gaming attracts a lot of new sales: not only is the firm offering lower fuel costs for the same price, it is also offering a price schedule unaffected by the regulation while all others comply. Similarly these two affects play against the firm when it is the only firm not gaming. This shows that if one firm decides to game other firms are under a lot of pressure to game along, especially when there is a binding policy.

6.2.4 Welfare loss from missing the externality target

If the regulator is unaware of gaming, or uncertain about its extent, the regulator will miss their targeted emissions reductions. The E.U. commission that regulates fuel consumption boasts about achieved savings, based on official ratings. Specifically, they claim that average emissions have fallen by 17g CO_2/km since 2010.[14] Our estimates suggest that about 75% of the reported decline in emissions from new vehicles is due to gaming. Thus, between 2010 and 2014, the actual decrease is 4.25 grams of CO_2/km. This translates to 32 million more tons of CO_2 emissions than the naïve estimate would suggest. At \$35 per ton, this equates to \$1.13 billion per year.[15] Analysts who perform cost-benefit analyses of the regulation may significantly mis-evaluate the program if they calculate savings based on official ratings.[16]

As discussed in section 2, the welfare losses from gaming are much broader than this. A naïve regulator will set the emission target at too high a level because they fail to include social costs of gaming. A sophisticated regulator will set a less ambitious (true) target when firms game than when firms are honest. In ongoing work we are setting up simulations to quantify the welfare losses under these scenarios (e.g., estimate the size of the Harberger triangles described by our theory). To do so we will need to make assumptions on the marginal cost of gaming relative to the marginal

[14] See for example `http://ec.europa.eu/clima/policies/transport/vehicles/cars/index_en.htm`.

[15] We assume that 12 million vehicles are sold across Europe, and that they are driven 14,000km per year for a 15-year period.

[16] This is for example the case in work of one of the current co-authors, see Reynaert (2015).

Table 8: Profit Changes from 5% gaming

	All Game (%)	Own Gaming (%)	All other game (%)
		No Policy	
Bmw	-0.00	16.15	-13.90
Daimler	1.18	16.74	-12.52
Fiat	1.09	17.93	-11.58
Ford	0.95	14.13	-11.27
General Motors	1.17	13.96	-10.93
Mitsubishi	2.02	16.13	-11.41
Nissan	2.05	16.47	-11.71
P.S.A.	0.55	13.45	-11.13
Renault	1.19	14.67	-11.29
Toyota	0.33	12.76	-10.91
Volkswagen	-0.14	11.02	-10.12
		With 5% Policy	
Bmw	-15.95	23.30	-31.68
Daimler	-11.77	28.68	-28.71
Fiat	13.14	39.86	-16.61
Ford	-4.96	21.44	-20.88
General Motors	4.71	28.24	-17.11
Mitsubishi	0.08	29.55	-21.27
Nissan	3.26	34.69	-22.00
P.S.A.	1.26	22.65	-16.82
Renault	0.09	25.15	-19.43
Toyota	0.97	27.70	-16.52
Volkswagen	0.13	19.37	-16.17

cost of actual investment in fuel consumption technology.

7 Conclusion

Our empirical analysis demonstrates a remarkable and growing divergence between official and on-road fuel consumption in Europe. Our analysis suggests that this discrepancy has significant implications for consumer surplus, which vary depending on the degree to which consumers were aware of the manipulation, and carbon mitigation. Regulators in Europe have recently become aware of these trends, and it remains to be seen how they will respond. Combined with the Volkswagen scandal related to local air pollution emissions from diesel vehicles, the facts we report on here suggest a veritable crisis in the administration of environmental regulations for automobiles.

Some insights on the implications for welfare and for policy design come from our theoretical model and contemplation of Goodhart's Law. We suspect that these lessons, and the cautionary tale of the Europe car market, are relevant for a variety of other regulations in the environment and beyond.

References

Allcott, Hunt. 2013. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." *American Economic Journal: Economic Policy* 5 (3):30–66. URL http://www.aeaweb.org/articles.php?doi=10.1257/pol.5.3.30.

Anderson, Soren T. and James M. Sallee. Forthcoming. "Designing Policies to Make Cars Greener: A Review of the Literature." *Annual Review of Resource Economics* .

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 64 (4):841–890.

Buchanan, James M. 1969. "External Diseconomies, Corrective Taxes, and Market Structure." *American Economic Review* 59 (1):174–177.

Chetty, Raj, John N. Friedman, and Johan E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9):2593–2632.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2016. "The Effects of Banning Advertising in Junk Food Markets." URL https://dl.dropboxusercontent.com/u/21832052/site/dgo.pdf. Manuscript: Toulouse School of Economics.

Duflo, Esther, Michael Greenstone, and Nicholas Ryan. 2013. "Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4):1499–1545.

Fowlie, Meredith. 2009. "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage." *American Economic Journal: Economic Policy* 1 (2).

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2015. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." Manuscript: UC Berkeley.

Fowlie, Meredith, Mar Reguant, and Stephen P. Ryan. 2016. "Market-Based Emissions Regulation and Industry Dynamics." *Journal of Political Economy* 124 (1):249–302.

Goodhart, Charles. 1981. *Problems of Monetary Management: The U.K. Experience*. Springer.

Goulder, Lawrence H. and Ian W. H. Parry. 2008. "Instrument Choice in Environmental Policy." *Review of Environmental Economics and Policy* 2 (2):152–174.

Grigolon, Laura, Mathias Reynaert, and Frank Verboven. 2014. "Consumer Valuation of Fuel Costs and the Effectiveness of Tax Policy: Evidence from the European Car Market." URL http://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxmcmFua3ZlcmJvfGd4OjZiMzQ4YzUwYjhiNjdjOTk. Manuscript: University of Leuven.

Ito, Koichiro and James M. Sallee. 2015. "The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards." Manuscript: University of Chicago.

Jacobsen, 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., Christopher R. Knittel, James M. Sallee, and Arthur A. van Benthem. 2016. "Sufficient Statistics for Imperfect Externality-Correcting Policies." Manuscript: University of California at Berkeley.

Jacobsen, Mark R. and Arthur A. van Benthem. 2015. "Vehicle Scrappage and Gasoline Policy." *American Economic Review* 105 (3):1312–1338.

Kane, Thomas J. and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper 14607.

Kartik, Navin. 2009. "Strategic Communication with Lying Costs." *Review of Economic Studies* 76 (4):1359–1395.

Langer, Ashley and Shaun McRae. 2014. "Step on It: Evidence on the Variation in On-Road Fuel Economy." Manuscript: University of Arizona.

Lucas, Jr., Robert E. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1:19–46.

Oliva, Paulina. 2015. "Environmental Regulations and Corruption: Automobile Emissions in Mexico City." *Journal of Political Economy* 123 (3):686–724.

Reynaert, Mathias. 2015. "Abatement Strategies and the Cost of Environmental Regulation: Emission Standards on the European Car Market." Manuscript: KU Leuven.

Sallee, James M. 2014. "Rational Inattention and Energy Efficiency." *Journal of Law and Economics* 57 (3):781–820.

Sigman, Hilary. 2012. "Monitoring and Enforcement of Climate Policy." In *The Design and Implementation of U.S. Climate Policy*, edited by Don Fullerton and Catherine Wolfram. University of Chicago Press, 213–225.

Sigman, Hilary and Howard F. Chang. 2011. "The Effect of Allowing Pollution Offsets with Imperfect Enforcement." *American Economic Review* 101 (3):268–272.

Small, Kenneth A. and Harvey S. Rosen. 1981. "Applied welfare economics of discrete choice models." *Econometrica* .

The International Council on Clean Transportation. 2014. "From Laboratory to Road: A 2014 Update of Official and "Real-World" Fuel Consumption and CO_2 Values for Passenger Vehicles in Europe." ICCT White Paper.

———. 2015. "From Laboratory to Road: A 2015 Update of Official and "Real-World" Fuel Consumption and CO_2 Values for Passenger Cars in Europe." ICCT White Paper.

Weyl, E. Glen and Michael Fabinger. 2013. "Pass-Through as an Economic Tool: Principles of Incidence under Imperfect Competition." *Journal of Political Economy* 121 (3):528–583.