

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*

November 21, 2016

Abstract

Firms sometimes comply with externality-correcting policies by gaming the measure that determines policy. We show theoretically that such gaming can *benefit* consumers, even when it induces them to make mistakes, because gaming leads to lower prices by reducing costs. We use our insights to quantify the welfare effect of gaming in fuel-consumption ratings for automobiles, which we show increased sharply following aggressive policy reforms. We estimate a structural model of the car market and derive empirical analogs of the price effects and choice distortions identified by theory. We find that price effects outweigh distortions; on net, consumers benefit from gaming.

Keywords: gaming, corrective taxation, environmental regulation, carbon emissions, automobiles, fuel economy

JEL: Q5, H2, L5

*We thank seminar participants at Arizona State, ENSAE-CREST, Fundação Getúlio Vargas, the NTA Annual Conference on Taxation, the NBER Summer Institute, the Stanford Institute for Theoretical Economics (SITE), 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 proclaims that “when a measure becomes a target, it ceases to be a good measure.”¹ It captures the notion that strategic actions will alter the meaning of a measure once stakes are introduced. 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 theoretically and empirically, with a focus on the case in which regulation is intended to mitigate an externality. We first develop our theoretical model, which characterizes the welfare effects of corrective policy in the presence of strategic gaming. We then apply these insights to the case of the regulation of carbon emissions from automobiles in Europe. We use a novel data set to show that the roll out of aggressive carbon policies coincided with a remarkable decline in the accuracy of laboratory-based carbon emissions ratings, which is the measure used by policy. This is Goodhart’s Law; policy alters the accuracy of the measure. To conduct welfare analysis, we estimate a structural model of the European car market and use the parameters to directly quantify the welfare effects identified in our theoretical model.

Our theoretical model considers a monopolist who sells a good to a representative consumer. The good has some attribute that is desirable to consumers, but also creates a negative externality that motivates corrective policy. The attribute is not directly observable, however, so consumer demand and government regulation are based on a signal sent by the firm. The firm can raise the signal either by changing the true attribute or by gaming, both of which are costly. Goodhart’s Law for externality-correcting policies manifests when the introduction of corrective policy induces gaming.

In our model, we assume that some fraction of gaming is undetected by consumers. Absent policy, this leads consumers to dislike gaming for two reasons. First, gaming causes consumers to mis-optimize (choose the wrong quantity of the good), which leads to a loss in consumer surplus that we call choice distortion.² Second, gaming causes the firm to raise price, because consumers perceive an improvement in the product. This further reduces consumer surplus. As a result, consumer incentives are misaligned with firm incentives regarding gaming; consumers lose from gaming and would shift demand away from firms that game if given the opportunity.

Corrective policy disrupts this logic by flipping the sign of price effects. Regulation raises the cost of production. Gaming allows the firm to lower its costs, and this benefits consumers through lower prices in the same way that a reduction in a tax would. When this price effect dominates choice distortions from faulty information, consumers *benefit* from the firm’s gaming, even when

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

²The choice distortion that we identify is conceptually identical to the consumer surplus loss due to cognitive frictions or information limitations in several prior papers, including Leggett (2002), Allcott (2013), Sallee (2014) and Dubois, Griffith, and O’Connell (2016).

they are fooled by it. Collectively, consumers would prefer to prevent gaming because the loss in private surplus would be outweighed by gains from mitigating the externality. But individually rational consumers wish to free-ride and would thus choose an environment with gaming to one with honest firms if given the choice.

Our theory articulates Goodhart’s Law for externality-correcting policies: stronger policy incentives lead to more gaming. This is intuitive, but less obvious are the welfare results we point out. Policy-induced gaming creates two welfare effects. First, firms spend real resources on gaming, which is socially wasteful. Second, gaming induces choice distortion. Combined with Goodhart’s Law, this means that strengthening corrective policy will exacerbate choice distortions and wasteful gaming, which together lower the social value created in the market. But, perhaps surprisingly, consumers may actually benefit (in terms of private welfare, not including the externality) from gaming, even when they are fooled by it, because of price effects. This means that consumers will prefer gamed products. In turn this suggests a challenge for enforcement. In most markets, consumer backlash is a powerful deterrent that keeps firms honest and provides market “self regulation.” But when consumers benefit from gaming, such self regulation will break down. We suggest that this break down of self regulation is precisely what happened in our empirical application.

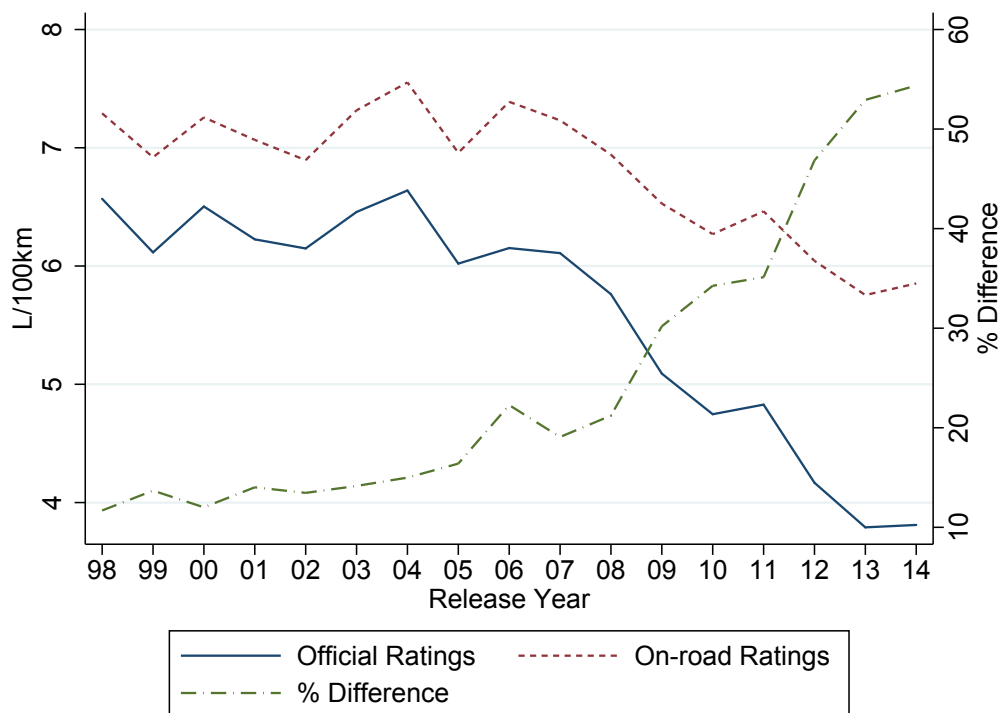
Our empirical analysis concerns a setting that closely matches our theoretical model: the regulation of carbon emissions from automobiles in Europe. In the automobile market, fuel economy is a characteristic of direct value to consumers, and it is directly linked to an externality (carbon emissions). In terms of the model, fuel economy ratings, which are isomorphic to carbon emissions ratings, are the true attribute. But, true on-road fuel economy is observed by neither consumers nor the regulator. Instead, regulations and consumer-facing fuel-economy labels are based on laboratory tests. Firms can improve a test result either by increasing a vehicle’s true fuel economy or by gaming the test.

Our empirical analysis begins by establishing a correlation between the accuracy of laboratory-based fuel-economy test ratings and changes in policy. Prior to 2007, there were no policies in Europe that hinged directly on fuel-economy tests. In 2007, the E.U. announced an aggressive Europe-wide fuel-economy mandate similar in structure to the U.S. Corporate Average Fuel Economy (CAFE) standard.³ The law was finalized in 2009 and has been phasing in since. In addition, a collection of nation-specific tax reforms changed, mostly between 2008 and 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 a common laboratory test called the New European Driving Cycle.

We compare these official test ratings to direct measures of on-road fuel economy constructed from a data set that tracks fuel consumption and kilometers traveled for a panel of more than 250,000 drivers over twelve years in the Netherlands. Using these data, we estimate the percentage difference between the laboratory test and on-road performance, which we call the *performance gap*. Figure 1 summarizes our findings; it plots the official rating, our estimated on-road performance,

³See [Reynaert \(2015\)](#) for an analysis of the economic effects of the European regulation.

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



and the corresponding performance gap by release year (model year) for vehicles in our sample. 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 50%. Below, we show that this result is robust to a number of specification checks and various controls, and that the performance gap is similar for all automakers. The rise in the performance gap implies that around 65% of the gains in fuel economy as measured by laboratory tests are false. Using conventional estimates of lifetime distance traveled and a social cost of carbon of \$40 per ton of carbon, the difference between apparent and actual emissions reductions amounts to \$1.2 billion annually from 2010 to 2014 when extrapolated to all of Europe.

Our results are consistent with a growing set of media accounts and with analysis performed by the International Council for Clean Transportation, which has documented a similar gap in a number of European countries (The International Council on Clean Transportation 2014, 2015).⁴ The media accounts offer insight into how gaming may occur.⁵ Tests are performed in third-party facilities, but these are funded by the automakers. The test procedure provides considerable latitude

⁴The ICCT reports include analysis of the Netherlands based on the same data that we use here. Our results are qualitatively similar, but we offer a number of critical methodological improvements, including the calculation of standard errors, the introduction of time of driving fixed effects and other controls, a variance decomposition, an Empirical Bayes correction, and extrapolation to the entire European market through matching.

⁵For example, see “Europe’s Auto Makers Keep Test Firms Close” in the March 21, 2016 *Wall Street Journal*: <http://on.wsj.com/1o5h47B>.

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 low-rolling resistance tires just for the test, optional equipment is removed (to reduce weight), and the outside of the body is altered (mirrors are removed, seams are taped down) to improve aerodynamics.

The historical data pattern is consistent with our theoretical model of Goodhart's Law. The same test procedure that is now being brazenly gamed was the basis of consumer-facing labels that are displayed by law on all new cars during our entire sample period. As such, as long as consumers cannot see through gaming perfectly, then automakers could have benefitted from test gaming during the early 2000's (prior to the roll out of policies), during which time fuel prices were quite high. Gaming in this time period, however, would have harmed consumers. Because the performance gap is observable over time (drivers observe their fuel economy), we would expect consumers to be fooled temporarily, but not permanently. Thus, gaming that harms consumers in one period may lead to a reduction in demand in future periods that acts as a self regulatory check on the market. The introduction of policy disrupts this logic. 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.

Our theoretical model shows that corrective policy creates price effects and choice distortions that have countervailing effects on consumer surplus, but it does not tell us in general which will dominate. To estimate the magnitude of these welfare effects and to see which dominates, we turn to structural analysis. We estimate a demand model of the European car market and use the structural parameters to simulate market outcomes from gaming with and without a policy. When there is no corrective policy and when consumers are completely unaware of gaming, we find that lowering perceived fuel costs by 5% reduces consumer welfare by €25 per car. A significant majority of this loss comes through the price effect rather than choice distortion. As a result, the profits of gaming firms rise by an amount similar to the loss in consumer surplus.

As suggested by the theory, the welfare effects of gaming change when we introduce a corrective policy. Specifically, we model a short-run regulation that requires firms to raise fuel economy by 5%. When firms comply honestly, private consumer surplus declines significantly. When we allow all firms to game, the constraints imposed by the regulation are relaxed, which creates a beneficial price effect for consumers. Gaming with or without a policy induces a similar sized choice distortion, but we find that this is an order of magnitude smaller than the price effect. In the presence of a policy, gaming causes private consumer surplus to *rise* by €200 per car. This provides empirical validation for our theoretical explanation: gaming benefits consumers. We also show that a corrective policy roughly triples the private benefit to a single firm that games when all others are honest. This shows that the incentives to game in the market are increased by the policy.

We contribute to several literatures. First, we make an important contribution to the emerging literature that assesses whether or not energy efficiency policies deliver the gains predicted by ex ante engineering-based procedures. [Fowlie, Greenstone, and Wolfram \(2015\)](#) show that building

insulation programs deliver a modest fraction of the expected benefits. [Davis, Fuchs, and Gertler \(2014\)](#) show a similar result for appliance rebate programs in Mexico. [Levinson \(2016\)](#) argues that building codes in California failed to produce the expected energy conservation. We contribute by showing similar results for the transportation sector; we estimate that two-thirds of the apparent gains based on laboratory tests are not realized. This literature is especially important for policy in light of the shift towards “bottom-up” contributions in international climate negotiations embodied in the 2015 Paris Climate Accord. In those proceedings, each country pledges expected emissions reductions based on ex ante assessments of future conservation programs. Energy efficiency policies play a large role in these estimates, and if the programs fail to produce the pledged gains, future rounds of negotiation will be greatly strained.

Second, by analyzing externality-correcting policies in a setting with market power, we contribute to the literature on instrument choice and policy design in the presence of multiple market failures. [Buchanan \(1969\)](#) argued against Pigouvian taxation on the grounds that such policies could lower welfare by exacerbating inefficiencies due to market power. Subsequent theoretical work demonstrated how to balance these two considerations ([Barnett 1980](#)), but relatively little empirical work followed. Recent empirical papers have begun to fill this void, including [Fowlie \(2009\)](#) and [Fowlie, Reguant, and Ryan \(2016\)](#), who use structural models of the electricity and cement markets in order to compare the efficiency of alternative policy instruments in the presence of market power and leakage. We add to this literature first by providing an empirical application to the automobile sector; second by introducing a new market failure, choice distortion due to gaming; and third by establishing an intuitive analytical model that shows how all three market failures interact.

Third, we contribute to the instrument choice literature. Reviews of this literature often cite ease of enforcement and compliance concerns as a reason to prefer one policy instrument over another, but cite little in the way of empirical evidence (e.g., [Goulder and Parry 2008](#)).⁶ We provide a key empirical lesson, and we also add to the instrument choice literature by establishing a novel theoretical point about how instrument choice can affect the ease of enforcement by aligning the interest of consumers either with firms that wish to evade or with the regulator. Specifically, we argue that gaming implies that a gasoline tax would have better enforcement properties than a fuel-economy regulation because a gasoline tax would align the interest of consumers with the regulator.

Finally, 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.

⁶Some recent papers have highlighted enforcement problems environmental regulation with a focus on developing countries (e.g., [Duflo, Greenstone, and Ryan 2013](#), [Oliva 2015](#)). We add to this literature by highlighting enforcement problems in a rich-country context, as well as through our theoretical insights.

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 model to describe the welfare consequences of strategic gaming in response to an externality-correcting policy. We consider the simplest case—a monopolist selling a single product to a consumer in a single period—that allows us to highlight the competing effects of consumer mistakes and endogenous prices. We focus our discussion on private surplus (excluding the externality) because our goal is to understand the compliance decision of individual firms and consumers, who would rationally free ride if given the chance.

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 attribute is not immediately verifiable, however, so consumers and the regulator must act on a signal provided by the firm. Deviation between the true attribute and the signal provided constitutes gaming.

Additional assumptions are motivated by the application we have in mind, where the attribute is energy efficiency of a durable good. 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. Both assumptions can be relaxed without altering our main conclusions.

Our notation is as follows. 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 L/100km rating). Note that x is a bad. Consumers and the regulator do not observe x directly, but instead receive a message m from firms (e.g., the fuel consumption 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 : $D(f) = D(p + \beta x)$. 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 \tilde{x} are assumed to be a weighted average of the truth and the signal:

$$\tilde{x}_t = \alpha x + (1 - \alpha)(x - g) = x - (1 - \alpha)g.$$

This is a tractable form of beliefs that encompasses a variety of possibilities. When $\alpha = 1$, consumers can see through gaming completely. When $\alpha = 0$, consumers are completely fooled. For intermediate cases, $(1 - \alpha)$ represents the fraction of gaming that fools consumers.

Policy intervention is motivated by a negative externality associated with x , which we assume is linear and equal to ϕx . The regulator observes m , so policy must be based on m . One possible policy is a regulatory standard, which requires that the (reported) attribute be below a threshold: 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 $\lambda \times (\sigma - x + g) \times D$.

Second, as an alternative, the regulator could raise the price of fuel consumption through a tax on fuel τ . We model this as a level increase in β , so that fuel costs with a tax are equal to $\tau + \beta$. 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 could add an additional cost function that is rising in the amount of gaming. But, rather than add a second term, we interpret $h(g)$, the cost of gaming function, as encompassing both private costs incurred directly by the firm to game as well as the expected value of penalties.

The monopolist chooses p , x and g to maximize profits, given the consumer demand system, the degree of consumer sophistication, and policy interventions. The firm's Lagrangean is:

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

Observe 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 $-c'(x) = \beta$.⁷ This says that they will

⁷The 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.

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 $\lambda = \phi$.

2.2 Absent policy, gaming lowers consumer surplus

Our main interest is in establishing how gaming affects consumer welfare in this setting. We start by showing how the firm alters its choice of p and x for any given level of g , which allows us to then characterize consumer surplus. The optimal degree of gaming will depend on the regulatory punishment function, which is embodied in $h(g)$, so this exercise can be thought of as examining how shifts in the punishment function that alter g would influence consumer surplus. We return to the endogenous choice of g in subsection 2.5, but for now we focus on characterizing how changes in g (which determine changes in p and x) influence consumer surplus.

Our first proposition shows that, when there are no corrective policies, an increase in gaming hurts consumers. Intuitively, 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 when there is no policy.

How does gaming affect x ? If there is no policy, then changes in gaming do not affect the firm’s choice of x . The profit-maximizing x satisfies $-c'(x) = \beta$, which is unaffected by gaming because the costs of x and g are assumed to be separable. As a result, the true attribute of the product will be unchanged when gaming is introduced, and private welfare losses come not from a misallocation of x but from misperceptions and market power.

How does gaming affect p ? 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 the same as the effect of *taxing* the *firm* by the same amount. 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 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 deadweight loss that is directly analogous to 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:

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

where the approximation assumes that the demand curve is locally linear. 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 properties common to Harberger triangles: it is zero when $g = 0$; it is rising with the square of the distortionary wedge; and it is larger when demand is more elastic.

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. [Proposition 1](#) expresses this formally.

Proposition 1. *In the absence of policy ($\sigma = \infty, \tau = 0$), consumer surplus falls with the level of gaming. Specifically:*

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

When consumers are fully sophisticated ($\alpha = 1$), gaming causes no change in consumer surplus; both terms go to zero. Both effects 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.

2.3 With a regulation, gaming raises consumer surplus

Now suppose that there is a binding regulation, but no tax on fuel ($\lambda > 0, \tau = 0$). Absent policy, the monopolist chooses x efficiently by setting $-c'(x) = \beta$. If there is no gaming, a binding regulation forces x below that private optimum. Lowering x improves the product for the consumer, but it raises costs for firms by even more than the improved private value to consumers. The difference is the shadow price of the regulation: $c' + \beta = \lambda$. As such, the regulation imposes a burden on the market, which will be shared between consumers and producers according to the standard logic of tax pass through.

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 implicit wedge λ . It is then obvious that the producer and the consumer *both* share in the benefits of gaming. Gaming lowers the regulatory 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 only when the benefits from the price effect 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 . An increase in g with a corresponding decrease in x holds constant reported fuel consumption costs, but 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 ($\alpha = 0$). Proposition 2 expresses this algebraically.

Proposition 2. *In the presence of a binding standard but no tax ($\lambda > 0, \tau = 0$), a change in*

gaming affects consumer surplus as follows:

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

Gaming does create a choice distortion here, just as in the situation without a policy. 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. The price effect, however, is changed significantly from the case with no policy, including a change in sign. As described above, the price effect can be positive. The next two corollaries describe cases where the $dCS/dg > 0$, so that consumers gain from gaming. Specifically, for a given level of gaming, dCS/dg will be positive for a sufficiently stringent standard, and, whenever the price effect is positive, dCS/dg is positive for a sufficiently small amount of gaming.

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 $\sigma > \sigma^*$.*

Intuitively, the wedge between marginal benefit and marginal cost of x is like a pre-existing distortion. Gaming exacerbates the choice distortion in order to reduce the wedge in the choice of x . Corollary 1 says that, when the wedge in x is large enough, the gains from mitigating that distortion will outweigh the losses from the choice distortion.

Similarly, because the choice distortion has the shape of a Harberger triangle, the initial distortion starting at $g = 0$ will be zero. This implies that, whenever the price effect is positive, consumers will benefit from the first unit of gaming. That is, consumers want at least some gaming whenever the price effect is positive. This is stated in corollary 2

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 and there is a binding policy, increases in gaming benefit 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 in our empirical application.

2.4 With a fuel tax, gaming lowers consumer surplus

When the planner uses the fuel price instrument instead of the regulation, the benefits of gaming no longer accrue to consumers in the same way. Conceptually, a tax raises the value of x to consumers. This leads the firm to provide more x , which will satisfy the first-order condition: $-c'(x) = \beta + \tau$. But, gaming does not help the consumer avoid the fuel tax τ , so it provides them no benefits. Unlike the case of a regulation, gaming under the fuel tax operates just like gaming in the case of no regulation. The only difference is that β has been raised by policy.

Proposition 1 showed that consumers lose from increases in gaming, for an arbitrary value of β , in the absence of a binding regulation. The situation with a fuel tax is exactly the same as that situation, but with an increased value of β . Thus, Proposition 1 implies that, when a fuel tax is used as the corrective policy instead of a regulation, the introduction of gaming causes a loss of private consumer surplus, both through a price effect and through choice distortion. Proposition 3 states this result.

Proposition 3. *In the presence of a tax but no standard ($\sigma = \infty, \tau > 0$), consumer surplus falls with the level of gaming. Specifically:*

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

Normally, the statutory incidence of a tax is irrelevant to its ultimate economic incidence; i.e., it does not matter if a retail sales tax is levied on buyers or sellers. But, this irrelevance need not hold when there are opportunities for avoiding or evading the tax. Our result here is a manifestation of this same result. In our model, a regulation σ creates a shadow price λ that corrects the externality associated with x . A tax τ does exactly the same thing. If there were no evasion, the two policies would be identical whenever $\tau = \lambda$, but gaming offers an opportunity to avoid a regulation but not a fuel tax, which breaks the symmetry and provides an enforcement rationale for preferring the tax over a regulation.

2.5 Goodhart’s Law: policy induces gaming

The above derivations considered how a change in the amount of gaming would influence consumer surplus, but it did not specify the profit-maximizing amount of gaming. The first-order condition for g , taken by differentiating equation 1, with a regulation (but no tax) is:

$$\underbrace{h'(g)}_{\text{MC}} = \underbrace{\lambda D}_{\text{Regulatory benefit}} + \underbrace{(p - c)}_{\text{Markup}} \times \underbrace{[-(1 - \alpha)\beta D']}_{\text{Demand shift}}. \quad (2)$$

This is not a closed form solution, but it provides clear intuition. The firm will increase gaming until the marginal cost from $h'(g)$ reaches the marginal benefit, which is a combination of the regulatory benefit, captured by the shadow price of the regulation, plus the increase profit due to a shift out in demand. Intuitively, a tighter policy (higher λ) raises the benefits of gaming and thus will lead to a higher g .

Note that when the consumer is sophisticated, $\alpha = 1$, the second term is zero. This is the case when consumer’s see through gaming completely. Then, a regulation induces the firm to game, where they would do zero gaming absent policy. A higher value of λ (a tighter policy) will increase gaming.

Our version of Goodhart’s Law is a comparative static result that says that gaming will rise when the policy tightens. When the regulator raises the stakes associated with the measure m , this gives rise to further gaming.

Proposition 4. *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, on net, as implied by Proposition 2. But, it does create a negative effect through choice distortion. And, 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 social costs to the degree that gaming requires real resources.

2.6 Policy implications and model extensions

Our model is deliberately simple so as to highlight as clearly as possible how the introduction of corrective policy can induce gaming, and how that gaming can have counterintuitive welfare impacts on consumers. Our main results are, first, that tighter corrective policy will induce more gaming, which incurs real resource costs from firms and induces choice distortion among consumers, and second, that gaming, even when it causes consumers to make mistakes, can lead to increases in private consumer surplus through price effects. What do these results imply for policy design?

Standard second-best reasoning will imply that optimal corrective policy when gaming is a concern will be attenuated away from the Pigouvian benchmark. When consumers can see through gaming completely, the regulator can tighten the standard sufficiently so as to actually achieve the desired change in x , but this will come at an added cost because gaming will nevertheless occur in equilibrium, and the costs embodied in $h(x)$ represent pure social waste. When consumers are at least partially fooled by gaming, then policy also creates choice distortion, which the planner will balance against externality gains.

An additional implication relates to self regulation of markets. In many markets, consumers play a key role in self regulating a market. When firms cheat, consumers shift demand away from that firm in the future. This consumer backlash is often a more potent check on firm behavior than the explicit penalties imposed by a regulator. But, when consumers benefit from cheating, they will not have the same incentive to regulate the market. Intuitively, consider a consumer with a choice between two scenarios, one where they know that gaming exists, but in which gaming still creates a choice distortion, and another where there is no gaming. When price effects outweigh choice distortion, individual consumers would choose the scenario with gaming. To formalize this, we could extend our model to consider two firms, one that can credibly signal its true x (be honest) and another that does not. When gaming helps consumers, consumers prefer the gaming firm, which captures the notion that market fails to self regulate. Honesty can only be sustained by a sufficiently strong punishment.⁸

Our model focused on the case of a monopolist. How would introducing competition affect our results more generally? [Weyl and Fabinger \(2013\)](#) show that the basic intuition of pass through applies to a very broad set of market structures, including the differentiated products competition that matches our empirical setting. As a result, our core results about how gaming can have positive price effects should hold quite generally. All that changes is that 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.

⁸Prior models of deception, like [Gabaix and Laibson \(2006\)](#) and [Heidhues, Kőszegi, and Murooka \(2016\)](#), allow the possibility that one firm could reveal the deception for all firms in the market. A key difference between those settings and ours is the role of the corrective policy, which creates a mechanism in our model for how deception can create private welfare gains for consumers.

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, measurement, and the potential for gaming

Our empirical application concerns carbon policies for automobiles in Europe, which include an E.U.-wide regulation, as well as national tax and subsidy schemes. The E.U.-wide regulation mandates that automakers sell vehicles that have a sales-weighted average carbon emissions rate below a certain level. Note that a vehicle's carbon emission rate is measured as a linear transformation (to convert units) of fuel consumption, which itself is just the inverse of fuel economy. Thus, while the E.U. policy is described as regulating carbon, it is directly analogous to the U.S. Corporate Average Fuel Economy standards, which are described as regulating fuel economy.⁹

Prior to 2007, there was no legally binding fuel-economy regulation in Europe. The standard was announced in 2007, passed into law in its final form in 2009, and phased in over several subsequent years. 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. Monetary standards for non-compliance are stiff. When the standard was announced, if they left their fleet unchanged, automakers would have faced an average fine of 1,250 euros per vehicle.

An automaker whose fleet is out of compliance can meet the standard by improving the fuel economy of their vehicle or by shifting the sales mix of their vehicles towards cleaner models. 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 deployment of additional fuel-saving devices or technologies ([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, with only modest changes in the composition of their fleet.

⁹To be precise, there are now parallel regulations in the U.S., one which regulates fuel economy that is administered by the National Highway Traffic and Safety Administration, and another that regulates carbon emissions that is administered by the Environmental Protection Agency.

On top of the EU standard all member states have separate taxation schemes for new vehicles. During the roll out of the E.U.-wide regulation, many member states adjusted their national policies so that tax schemes were based explicitly on carbon emissions. France introduced a feebate, which taxes heavily-polluting cars and subsidizes cleaner models, in 2008. In 2009, Germany switched its annual road taxes so that they depend on carbon emissions rates, rather than engine cylinders. In 2008, Spain introduced registration taxes that depend on emissions ratings. The Netherlands did the same in 2010.¹⁰

Carbon emissions ratings for all of these policies are based on a laboratory test, called the New European Driving Cycle (NEDC). This test procedure, which is conducted by third-party facilities that are funded by the automakers, 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 city and highway driving. A coast down test is also performed to measure aerodynamics.

The NEDC is not only the basis of CO₂ 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 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 was used as a regulatory and market instrument for roughly a decade before the carbon regulation 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 not expressly prohibited from submitting test vehicles that have been modified. Firms remove optional equipment, thereby changing the weight of the vehicle to improve performance. They also tape down seams in the vehicle, remove side mirrors and roof racks, and over-inflate tire to improve results of 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.¹¹ To detect gaming, we compare laboratory test ratings with a direct measure of on-road fuel consumption, which we describe next.

¹⁰See the ACEA tax reports (www.acea.be) for an overview of all vehicle taxation in the EU and [Gerlagh, van den Bijgaart, Nijland, and Michielsen \(Forthcoming\)](#) for an overview of the change towards carbon taxation in EU member states after 2007.

¹¹These “flexibilities” differ significantly from analogous tests in the U.S., where details about the test vehicle and test procedures are described in much more detail by the regulation.

4 Measuring on-road carbon emissions

Our goal is to compare official test results and 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. For this task, 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. We then estimate how on-road performance changes with vehicle vintage and relate that pattern to the roll out of policy.

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 fuel. 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. Nearly half of new vehicles sold in the Netherlands are registered as company cars (Booz & Company 2012).

When visiting a fuel station, Travelcard NV users 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. Regarding the latter, drivers are asked to enter their current odometer reading into the fuel pump's keypad during the transaction. Transactions are linked to an account, which is a unique combination of a driver with a particular vehicle, which corresponds to a unique license plate. Unfortunately, if the same individual drives one vehicle for some time and then switches to another vehicle, we have no way of linking those data and must treat them as separate accounts. We refer to this panel variable as a driver, which should be understood as a unique combination of a driver with a particular vehicle.

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 volume, list price and a broader set of characteristics, including length, width and several 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 24 million transactions. In robustness tests, we find that censoring our data less severe only impacts our estimate of the level of the performance gap, not of the change in the performance gap over time.

Our data selection process first eliminates a number of drivers (accounts) that have unreliable information or too few observations. With the accounts that remain, we then consider several ways to account for mismeasurement in odometer readings for individual transactions. Specifically, we first limit the sample to gasoline- or diesel-fueled vehicles, which eliminates 6.7 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).¹² We drop car models with fewer than 10 drivers, and drivers with fewer than 10 fuel station visits (1.3 million transactions). We drop drivers that did not report driving more than 5,000km in total or reported driving more than 500,000km in total (11.3 million). Having isolated a set of drivers (accounts) with ample data, within those accounts we drop individual transactions in two steps. First, we drop transactions where the odometer difference is lower than 100km or higher than 3000km (7.6 million). Second, we drop transactions 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 (5.1 million). This results in the final dataset of 24 million observations.

The final data includes over 2,500 unique types of cars driven by 266,000 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 L/100km and a weight of 1,354kg. Somewhat less than half of the vehicles (46%) have diesel engines. We observe an average of 107 drivers per car with a maximum of 3,228 drivers. For each of these drivers we observe an average of 134 visits to the pump with a maximum of 1,135. Driver mean total consumption is 6,015 liters of fuel purchased corresponding to 111,726 km travelled. Finally, the average fuel station visit involves 45 liters of fuel purchased, corresponding to an odometer increase of 671 km with a standard deviation of 192 km.

Our sample is drawn from company cars, but our data provide estimates that characterize a majority of the market. In appendix Table A.1, we show that vehicles in the Travelcard data are cheaper, lighter, more fuel efficient and more likely to be a diesel than the average new vehicle in the Netherlands. One explanation of the difference is that our data contain almost no luxury vehicles or sports cars. Also, company cars are driven more than the average household car, which likely explains the higher share of diesels in our sample.

Nevertheless, the models in our final sample are the models that account for 76% of all sales in the Netherlands. Moreover, the Dutch car market is very similar to the entire EU market; it features a mix of French, German and foreign brands that is very similar to the E.U. average. Thus, our

¹²Note that the range of a combustion engine is easily more than 800km.

Table 1: Summary statistics

	Mean	St. Dev.
Car Characteristics		
Fuel Consumption (L/100km)	6.65	1.73
Vehicle Weight in kg	1,354	245
Diesel Engines	0.46	0.50
Drivers per car	107	219
Driver Characteristics		
Pump visits	134	80
Total liters purchased	6,015	3,666
Total distance (km)	111,726	53,942
Pump Visit Characteristics		
Liters per visit	45.3	10.8
Odometer increase per visit	671	192

The table gives summary statistics for the 2,696 vehicles, 266,616 drivers and 23,989,576 pump visits in the sample.

data provide estimates of on-road performance that cover the broader market fairly well. However, we wish to stress that we have no way of directly assessing whether drivers of company cars drive their vehicle differently, and thereby exhibit a different performance gap, than the average Dutch driver. Fortunately, we are focused on *changes* in the performance gap over time, so as long as differences between company car drivers and other drivers are stable over time, this will be less of a concern for our conclusions.

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:

$$r_{nij} = \frac{\text{liters}_n}{\text{odometer}_n - \text{odometer}_{n-1}} * 100. \quad (3)$$

This measure of on-road fuel consumption, in units of L/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 distance traveled, 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:

$$d_{nij} = \frac{r_{nij} - li_j}{li_j}, \quad (4)$$

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 .¹³ In the end, both approaches yield similar estimates. In a first exploratory approach we estimate regression equations on the microdata 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. The regressions 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. For example, we observe both a 2008 and 2009 vehicle type being driving in 2010, and so we can estimate vintage effects (2008 versus 2009) while controlling for time of driving (2010).

To study potential heterogeneity in gaming at the vehicle level we need vehicle level estimates of the performance gap. This motivates our second approach. In obtaining these estimates we want to take into account the underlying data quality for each vehicle controlling for the large variance in reported odometer readings, and the variation in the number of drivers and visits observed for different types of cars. Therefore, we follow the teacher value added literature, specifically [Chetty, Friedman, and Rockoff \(2014\)](#) and [Kane and Staiger \(2008\)](#), to estimate on-road fuel consumption r_j using precision weights and an empirical Bayes correction. 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.¹⁴

We start by decomposing the total variance in the sample $Var(r_{nij}) = \sigma_r^2$ into 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 pump visits of the same driver as:

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

in which \bar{r}_{ij} is the mean fuel consumption of driver i , N is the total number of observations and I is

¹³We have also worked with a third approach which is to take long differences between last and first visit of each driver. Estimates of the gap computed with long differences are reported in the appendix. Taking long differences might be more robust to bias from reporting errors but does not allow us to control for time of driving effects in the regressions. Results are very similar across approaches

¹⁴Note that this approach does not allow us to control for selection issues where drivers with different performance gaps select into different type of vehicles. Correlations between driving style and vehicle selection could influence our results, but again we emphasize our focus on changes in the performance gap over time, which mitigates these concerns.

the total number of drivers. Next, we estimate the covariance between drivers of the same vehicles as:

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

The estimated covariance is obtained as a weighted average of covariances between randomly sorted pairs (i, k) of drivers of the same car. We weigh each pair of drivers (i, k) by the sum of their visits. 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 \bar{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 \hat{r}_j to construct an alternative estimate of the gap defined in (4) and to inform us about the distribution of the gap between vehicles.

5 Estimates of the degree of gaming

5.1 Mean fuel consumption and gap

Figure 1, which we discussed in the introduction, plots the mean official rating li_j , the mean on-road rating r_{nij} and the percentage-gap d_{nij} for each release year. Between 1998 and 2006 we see that both official fuel consumption and on-road consumption vary between 6.1 and 7.5 L/100km. The percentage gap remains steady, between 12%-20%. From 2006 onward we see a spectacular drop in official consumption from 6.2 to less than 4 L/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 5.9 L/100km. As a result, the percentage gap between the official and on-road ratings increases dramatically: from around 12% at the beginning of the sample to almost 55% by 2014. This divergence is remarkable in magnitude, and it coincides exactly with the new EU-wide fuel consumption regulation, which was announced in 2007, finalized in 2009, and phased-in over the remainder of the sample, as well as the phase-in of national policies described in Section 3.

The divergence in test ratings and on-road performance is not isolated to a particular automaker.

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

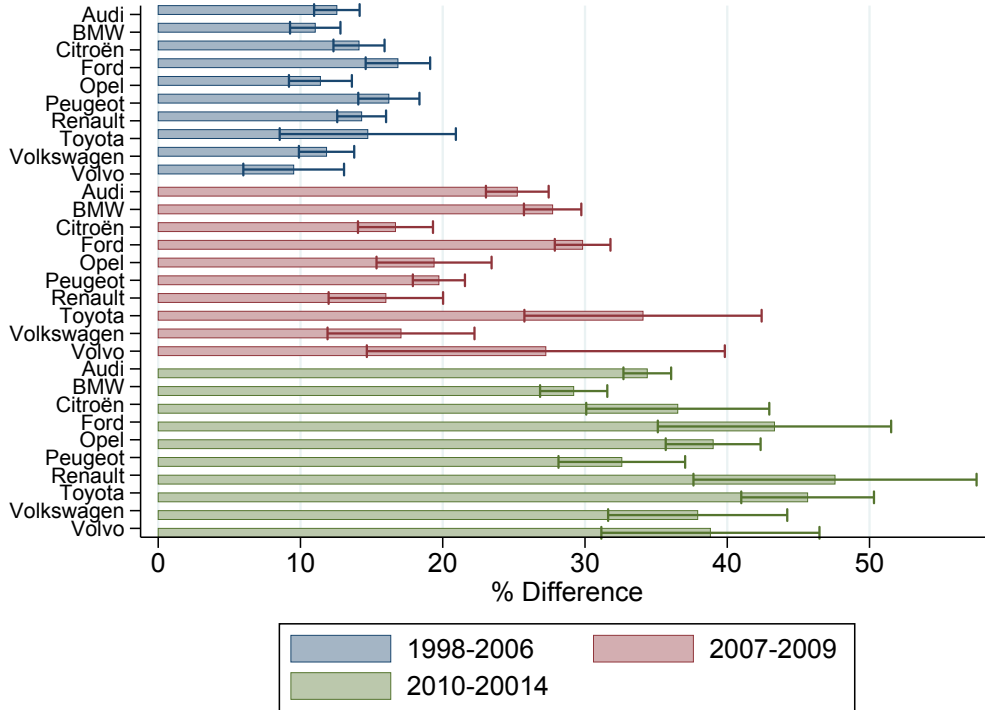


Figure shows estimated coefficients and standard errors from regressing the performance gap on three sets of model release years (early, middle and late) per automaker, the data is restricted to the ten brands with most observations.

Figure 2 plots the estimated mean gap d_{nij} for three sets of model years (early, middle and late) separately for each automaker. All automakers show a substantial increase in the gap over time, and all show an economically important gap in the later years. The performance gap is a global phenomenon: the three largest gaps are for a European firm (Renault), a Japanese firm (Toyota), and an American firm (Ford).

5.2 Fixed Effects Regression

Figure 3 plots the coefficients and their 95% confidence intervals (with clustering on vehicle type j) from two regressions with individual refueling transactions as the unit of observation. In the first regression (red markers) we estimate the percentage-gap d_{nij} as the coefficients on 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 before.

The second regression (blue markers) adds controls for time of driving and model fixed effects. Specifically, we add year and month of driving dummies to control for driving conditions that change over time.¹⁵ We add fuel-type (e.g., diesel) and model fixed-effects to control for compositional

¹⁵We observe later year vintages disproportionately in later years, so changes in congestion or other driving patterns

changes in the car market over time. If some vehicles have a bigger gap than others, then our time trend could reflect compositional changes. Controlling for model, which is more aggregate than our vehicle type j (e.g., a Toyota Camry is a model, whereas a 2010 Toyota Camry with 2.0L engine is a vehicle type), isolates variation between different engine releases over time of the same model.

Figure 3 shows that the introduction of controls has little impact on our results, which suggests that compositional changes or changes in driving conditions do not explain the time trend. The same coefficients are presented in columns (1) and (2) of Table 2.¹⁶

Our interpretation of the time trend is that policy caused a rise in gaming. What else might explain the 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 separately for gasoline and diesel engines in columns (3) and (4) of Table 2. There is initially no difference in the gap between diesels and gasolines, and this changes substantially over time. But, the coefficient patterns do not point to a 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 Distribution of the vehicle level gap

In this subsection, we consider the variation in the performance gap across drivers and across vehicle types, which serves two main purposes. First, our procedure provides an alternative way to estimate the performance gap that employs an empirical Bayes correction to account for variation in precision across drivers and vehicle types. Second, we show that variation in the performance gap has been relatively constant, which informs our subsequent welfare analysis. Specifically, in our final section we estimate a structural model of the automobile market and simulate counterfactuals. We choose to model a mean shift in the performance gap because of our finding that heterogeneity

could potentially bias our results if we omit time of driving controls.

¹⁶Note that the R^2 for the model without controls is .28, which suggests that release years explain an impressive fraction of the variation, given that the unit of observation here is an individual fuel transaction.

Figure 3: Release year coefficients from fixed effect regressions

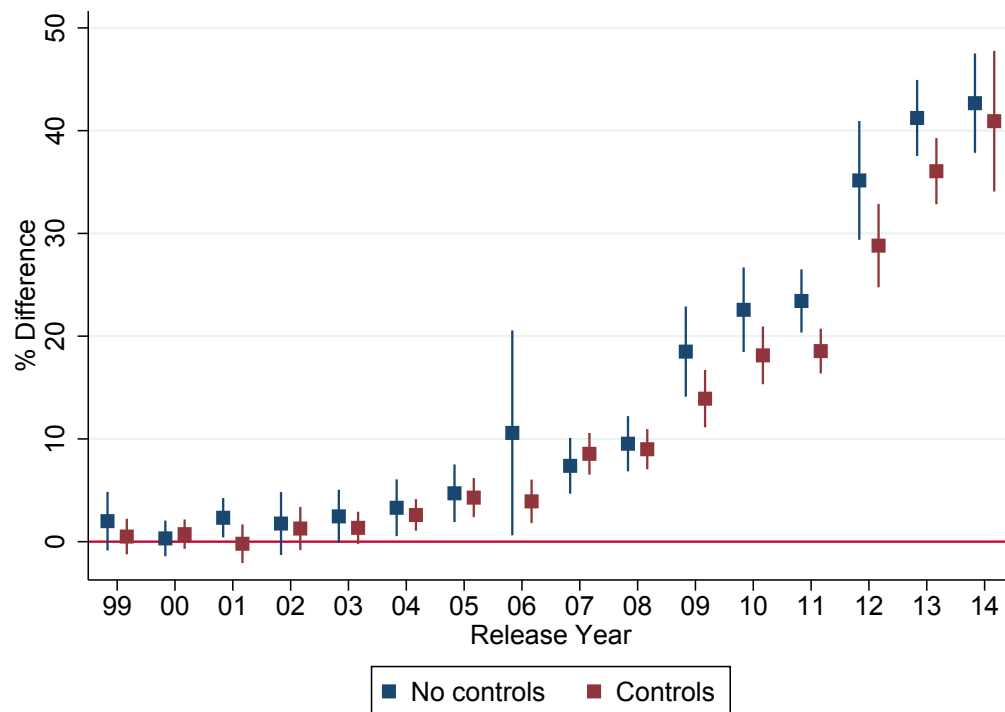


Figure plots coefficients from a regression of the performance gap (d_{nij}) on release year fixed effects. Coefficients correspond to regressions (1) and (2) from Table 2.

Table 2: Release year fixed effects from regression of performance gap on controls

	All (1)	Controls (2)	Gasoline (3)	Diesel (4)	1st Year (5)
Constant	11.69 (0.66)	13.59 (0.60)	11.44 (0.67)	11.94 (1.13)	10.28 (1.39)
1999	2.00 (1.45)	0.50 (0.88)	0.96 (1.10)	2.33 (2.10)	
2000	0.32 (0.88)	0.73 (0.73)	0.13 (1.21)	0.40 (1.33)	
2001	2.33 (0.98)	-0.20 (0.96)	2.08 (1.16)	2.40 (1.50)	
2002	1.77 (1.56)	1.28 (1.07)	2.40 (1.50)	1.33 (2.32)	
2003	2.46 (1.32)	1.35 (0.80)	3.79 (1.60)	1.16 (2.03)	
2004	3.30 (1.41)	2.60 (0.78)	3.93 (2.69)	2.80 (1.56)	
2005	4.71 (1.43)	4.29 (0.97)	8.55 (1.84)	2.54 (1.99)	2.50 (2.39)
2006	10.59 (5.09)	3.93 (1.08)	14.71 (5.96)	3.40 (1.67)	5.80 (1.73)
2007	7.38 (1.39)	8.55 (1.04)	10.75 (1.50)	5.02 (1.92)	2.79 (2.09)
2008	9.53 (1.37)	9.00 (1.00)	10.78 (1.09)	8.53 (2.33)	12.60 (2.47)
2009	18.50 (2.24)	13.92 (1.42)	19.14 (2.94)	17.61 (3.18)	15.66 (2.21)
2010	22.57 (2.10)	18.13 (1.43)	16.85 (1.42)	25.36 (2.80)	19.85 (2.09)
2011	23.43 (1.57)	18.54 (1.11)	23.10 (3.64)	23.54 (1.42)	23.28 (2.10)
2012	35.16 (2.95)	28.81 (2.07)	27.51 (2.17)	40.81 (2.52)	36.96 (4.38)
2013	41.24 (1.89)	36.06 (1.64)	35.26 (1.97)	43.38 (1.90)	43.16 (2.48)
2014	42.68 (2.47)	40.92 (3.49)	39.98 (1.69)	48.99 (2.46)	41.50 (2.52)
Year/Month F.E.		Yes			
Fueltype		Yes			
Model F.E.		Yes			
#Obs. (*10 ⁶)	23.98	23.98	10.17	13.81	0.91
R ²	0.28	0.40	0.21	0.36	0.26

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. Standard errors clustered by car type in parentheses. 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 3: Variance decomposition

	σ_r^2	σ_i^2	σ_j^2	σ_n^2
Mean	2.11	0.21	1.35	0.56
Standard deviation	0.57	0.03	0.45	0.11
Variance decomposition (%)	100	10.36	62.34	27.30
Standard deviation		2.54	6.76	4.40

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 co-variance between drivers in the same vehicle, 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.

is stable.

Table 3 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). We decompose the variation separately for each release year and describe the mean and standard deviation across release years in the table.

More than 25% 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 0.21. This is an economically large number; it means that the on-road fuel consumption is estimated to be 0.28 liter/100km higher at the third quartile than at the first quartile of drivers in the same car.¹⁷ 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 3%. These numbers are interesting from a policy perspective as they give an indication of the extent to which fuel consumption and emissions can be reduced by teaching and incentivizing drivers to drive a vehicle more efficiently.¹⁸ 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 1.35 or above 60% of the total variance. Table 3 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.

¹⁷If we assume that conditional on car j , r has a normal distribution, the interquartile distance is $1.349 \cdot \sigma_i^2$.

¹⁸Significant variation across drivers of identical cars is consistent with results reported in Langer and McRae (2014), who analyze extremely detailed driving data from a few dozen drivers of an identical car, the Honda Accord. In contrast, our data come from a large sample and cover many models.

Table 4: Estimated vehicle gaps by release year, with and without empirical Bayes correction

Release	Unweighted r_j	\hat{r}_j	Shrinkage	\hat{d}_j
1998	7.91 (1.41)	7.87 (1.39)	0.99 (0.00)	7.85 (7.82)
1999	7.90 (1.49)	7.88 (1.46)	0.99 (0.01)	8.70 (6.90)
2000	7.96 (1.44)	7.94 (1.43)	0.99 (0.01)	8.60 (7.36)
2001	7.73 (1.44)	7.72 (1.41)	0.99 (0.00)	11.15 (7.12)
2002	7.63 (1.58)	7.60 (1.57)	0.99 (0.00)	11.02 (9.17)
2003	7.83 (1.61)	7.80 (1.56)	0.99 (0.00)	11.01 (9.18)
2004	8.31 (1.68)	8.29 (1.64)	0.99 (0.00)	9.52 (9.58)
2005	7.87 (1.71)	7.82 (1.59)	0.99 (0.01)	12.56 (9.14)
2006	8.10 (1.57)	8.06 (1.53)	0.99 (0.00)	12.55 (8.17)
2007	7.71 (1.37)	7.68 (1.34)	0.99 (0.00)	17.66 (9.67)
2008	7.46 (1.36)	7.45 (1.34)	0.99 (0.00)	17.57 (9.00)
2009	7.24 (1.30)	7.21 (1.28)	0.99 (0.00)	22.34 (10.07)
2010	7.15 (1.33)	7.13 (1.32)	0.99 (0.00)	26.38 (10.68)
2011	6.92 (1.24)	6.90 (1.21)	0.99 (0.01)	29.74 (10.29)
2012	6.64 (1.08)	6.62 (1.06)	0.99 (0.01)	37.46 (11.05)
2013	6.23 (0.92)	6.23 (0.89)	0.99 (0.01)	44.23 (11.34)
2014	6.18 (1.20)	6.15 (1.13)	0.98 (0.02)	52.90 (12.04)

Table reports mean and standard deviations for the distribution of estimated on-road consumption r_j , Bayes corrected on-road consumption \hat{r}_j , shrinkage factor and Bayes corrected efficiency gap \hat{d}_j by release year.

Figure 4: Distribution of vehicle level fixed effects

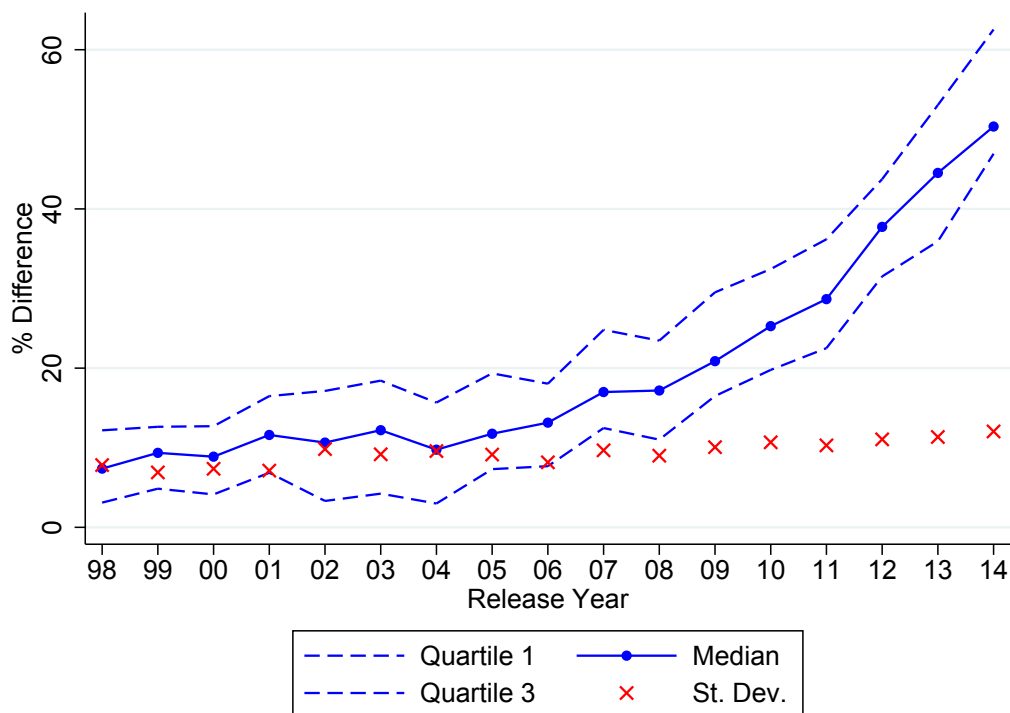


Table 4 reports the unweighted mean estimate of r_j and \hat{r}_j , obtained with the empirical Bayes correction. The mean value of both r_j and \hat{r}_j are decreasing over the release years. In all years the corrected means are lower than the raw means, because on average vehicles with high r_j have less precise underlying data, but overall shrinkage and precision weighting has no significant effect. The resulting gap \hat{d}_j is estimated to be an imprecise 10% up until 2006. From 2007 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 7% and 12%. To see this, Figure 4 plots the quartiles of the estimated gaps for all release years as well as the standard deviation of the estimated fixed effects. Again we see a clear increase in the estimated gap over time but somewhat surprisingly this increase in the gap is not associated with a considerable changes in the distribution of the gap. If firms would game only on some vehicles but not on others we would see a significant increase in the spread between the quartiles and the standard deviation. However, the standard deviation of the estimated fixed effects is constant and does not reveal this pattern.

To perform our simulations, we need to decide how to model gaming. Should we model gaming as a shift in the performance gap common to all vehicles, or should we try to assign different levels of gaming to different models? With our data, it is difficult to assign heterogeneity in the performance gap to heterogeneity in gaming. Some of the heterogeneity in the performance gap comes from selection of different types of drivers, who drive different routes on average and who

may be more or less aggressive in driving style, into different models. when we regress the gap on vehicle characteristics we find evidence supporting this: more fuel efficient vehicles have higher gaps, consistent with aggressive drivers sorting into more fuel efficient vehicles. In later years, however, we are not able to disentangle selection from heterogeneity in gaming. Given that we find a constant spread in the gap between vehicles, the assumption that selection is constant over time would be sufficient to conclude that gaming is an industry wide phenomenon. This motivates our main assumption for the simulations, which is that changes in gaming should be modeled as a common level of gaming across products, though in the appendix we consider simulations that introduce heterogeneity in gaming.

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 10% to 15% before regulation. By 2012 this gap has increased to 30% and all estimates give a gap of more than 45% in 2014. We do not find a single robustness check where this spectacular increase in the gap does not occur. Our aggregate findings are consistent with industry analysis performed by the International Council for Clean Transportation ([The International Council on Clean Transportation 2014, 2015](#)).¹⁹ 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 industry wide gaming for the evaluation of environmental policy.

6 Quantifying the effects of gaming

To quantify the welfare effects of gaming, we first estimate a demand model for new vehicles in order to obtain taste parameters for consumers. Given these estimated taste parameters, we simulate counterfactuals and calculate changes in consumer surplus that result from gaming, under alternative assumptions about the policy environment. Our simulations are informed by our reduced form results and by our theory. Specifically, we show that there is a direct analog to the choice distortion and price effects identified by our theory that can be decomposed from the standard log-sum welfare formula for discrete choice models. And, we model a mean shift in the distribution of gaming, consistent with our reduced form results. We proceed by describing the estimation, simulation and computation of consumer surplus and conclude with the results of the counterfactual. Throughout, our focus is on private consumer surplus. In our last step, we return to the externality and consider the welfare consequences of gaming related to externality mitigation.

¹⁹The ICCT reports results from various sources including the data from Travelcard NV we obtained. The other sources are a consumer reporting website and consumer car testing magazine. The data from Travelcard are by far the most extensive in terms of number of vehicles and drivers included, but it is reassuring that all data in the ICCT reports point to very similar results.

6.1 Estimation of the automobile demand system

We begin by modeling the choice of a consumer making a discrete choice about which vehicle to buy. 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 euros per kilometer). Utility from the outside good (not buying a new vehicle) is normalized to zero, $u_{i0} = 0$. We estimate a random coefficient logit model and assume that β_i and γ_i are independently normally distributed. We estimate the mean and standard deviation of β_i and γ_i . All remaining consumer heterogeneity is contained in the additive idiosyncratic error term ε_{ij} . The ξ_j term represents the value of product attributes unobserved by the researcher but observed by firms and consumers.

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 a panel containing sales, prices and characteristics for all new vehicle sales in seven European countries between 1998 and 2007.²⁰ We only use data from before 2008 to estimate demand, so that our estimates come from a period in which the performance gap was stable. We thus assume variation in x_j is informative about actual fuel cost differences. The vector Δ_j contains information on horsepower, weight, footprint (a measure of vehicle size), height and a dummy specifying if the car is of a foreign brand (e.g., Fiat in France). Additionally, we include fuel type by market dummies, dummies for the number of months a vehicle was on sale in a country-year, country fixed effects, a linear time trend, body type fixed effects, vehicle class fixed effects and brand fixed effects. We divide prices by income per capita in each country-year, so that price sensitivity varies with income in the market. We instrument for prices using both cost shifters and sums of characteristics instruments. The sums of characteristics instruments are the sum of fuel costs, horsepower, weight, footprint and height across all other products on the market and across all other products within the same firm on the market. We also include the number of competing products and the number of products in the same firm. The cost shifters are the log of labor costs in the country of production and a dummy specifying if the vehicle is sold in the country of production. Approximate optimal instruments for the standard deviations of the taste distributions are constructed using a two step procedure as described in [Reynaert and Verboven \(2014\)](#).

²⁰The seven countries are: Belgium, France, Germany, Great Britain, Italy, Netherlands and Spain

Table 5: Estimation Results

	Mean Taste		St. Dev.	
	Coeff.	St. Error	Coeff.	St. Error
Price/Inc.	-6.51	(0.45)		
Fuel Cost	-0.53	(0.01)	0.05	(0.03)
Horsepower	1.48	(0.21)	1.78	(0.10)
Weight	0.22	(0.21)	4.32	(0.16)
Footprint	0.88	(0.05)	0.58	(0.04)
Foreign	-0.92	(0.03)	0.02	(0.16)
Height	0.02	(0.02)		
Doors	0.50	(0.11)		

The table shows estimated taste parameters from a random coefficient logit estimation on the the car market for seven EU countries using data from 1998 to 2007. Taste distributions are assumed to be normal, and mean and standard deviations are estimated for selected characteristics. Additional controls are fuel type by market dummies, months for sale if less than 12, country fixed effects, linear time trend, body type fixed effects, vehicle class fixed effects and brand fixed effects. Model is estimated using a two-step GMM using approximate optimal instruments with sum of characteristics and cost shifter instruments for prices.

Estimated taste parameters and standard errors are reported in Table 5. Price and fuel costs have the expected negative effect on utility. The mean of the estimated own price elasticities is -5.45, in line with the previous literature. We estimate considerable heterogeneity in the taste for horsepower, weight and footprint. Vehicles perceived as foreign are less attractive for consumers. Cars with four doors are preferred over cars with two doors. We next use these estimates to study market outcomes when firms game.

6.2 Counterfactual market outcomes from gaming

Simulation setup: Our theoretical model establishes how gaming affects consumer welfare, emphasizing how the presence of a corrective policy interacts with gaming. In this section, we use our structural estimates to quantify changes in consumer surplus due to gaming under alternative assumptions about policy. Our theory focused on the case of monopoly, while the simulations consider a differentiated product oligopoly. We show that direct analogs to our theoretical objects are available in a decomposition of the log-sum formula for this case.

In all simulations, we start with a base market, which is the observed market in the Netherlands in 2007. We compute consumer surplus from this observed market and our parameter estimates. We assume that firms can game by 5%, which is roughly consistent with the yearly change in gaming estimated in our data. To model gaming, we first lower the fuel cost of firms that game by 5%. Second, with these new fuel costs, we solve for demand and prices using the first-order conditions for profit maximization. Third, we compute changes in welfare relative to the base market.

We model the policy as a requirement that each firm decrease its sales-weighted emissions by

5%.²¹ Absent gaming, firms must comply by changing prices to shift their sales to efficient models in order to become compliant. When we introduce gaming, firms that game can meet the policy requirement without this sales shift. Note, however, that the market equilibrium will nevertheless change to the extent that consumer demand, and hence prices and quantities, shift in response to gaming.

Our procedure assumes that gaming is a fixed cost: it does not shift the marginal costs of the firms. This is in line with evidence on how firms are gaming the test in Europe. Our procedure also assumes that consumers do not see through gaming and thus base their choices on the stated fuel costs at the time of purchase. In each scenario we discuss what happens if consumers are sophisticated, and the procedure could be scaled to reflect any ratio of consumer sophistication (α , in terms of the theoretical model). The obtained changes in welfare are the changes in yearly utility, profits and emissions from new vehicle sales. Next, we describe how we compute consumer and producer surplus.²²

Computation of consumer surplus and profits: To compute consumer welfare when consumers are affected by gaming ($\alpha < 1$, in terms of our theory) we follow [Dubois, Griffith, and O’Connell \(2016\)](#), who describe the welfare effects of persuasive advertising. As described in [Section 2](#), gaming changes prices from p_j^0 to p_j^1 , as firms will exploit higher demand from lower perceived fuel costs. This misperception of fuel costs also distorts consumer choice. To separate the two effects we make a distinction between decision (at the moment of purchase) and experience utility (at the moment of utilization).²³ When there is gaming, a naïve consumer will perceive fuel costs, following her belief, as $\tilde{x}_j = d_j * x_j$, in which d_j is the increase in gaming. The consumer will make her choice based on \tilde{x}_j and will perceive her decision as yielding a utility of:

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

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

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

²¹We use the actual formula used by the EU in its emission standard, so that the policy is based on vehicle weight. For each firm we compute a sales weighted average emission, with emission for heavier (lighter) cars receiving a bonus (penalty) in the weighted sum. See [Reynaert \(2015\)](#) for a detailed description.

²²Note that we limit ourselves to reporting changes in consumer welfare from purchasing a new vehicle. We do not consider effects on the second hand market or on the amount that consumers drive. Consumers could potentially react to gaming by changing the amount they choose to drive in the purchased vehicle.

²³This is conceptually the same as the choice distortion described in [Allcott \(2013\)](#) and [Sallee \(2014\)](#) who study misperception of fuel costs, but our approach here is more general because we allow firms to change prices in response to gaming. [Leggett \(2002\)](#) also models a similar distortion in a discrete choice setting. Note that we implicitly assume that consumers have on average a correct belief about future fuel costs before gaming is introduced. If there were undervaluation of fuel costs, gaming could have a third effect in mitigating externality costs as described in [Allcott and Wozny \(2014\)](#). However, several papers find that undervaluation is small or non-existent. See [Busse, Knittel, and Zettelmeyer \(2013\)](#), [Sallee, West, and Fan \(2016\)](#) and [Grigolon, Reynaert, and Verboven \(2014\)](#), the latter of which studies the EU market.

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 $\beta_i(x_j - \tilde{x}_j) = \beta_i(1 - d_j)x_j$. 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 honest 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 markup and demand, both of which are a function of the price vector. Policy forces a non-gaming firm to change its price schedule, while gaming shifts the demand and also changes the optimal prices chosen by firms.

Market outcomes from industry wide gaming: Results describing the market effects of gaming are given in [Table 6](#). The column on the left shows the effects on consumer surplus, pollution and profits when there is no binding policy. The column on the right gives the effects of gaming with a 5% policy in place.

Absent policy, total consumer surplus per household decreases by €24 when gaming is introduced. The choice distortion is responsible for €17 of the decrease. Because fuel consumption ratings are shaded by the gaming, consumers purchase the wrong amount and the wrong type of vehicle. The remaining decrease in consumer welfare, €7, comes from the price distortion. Consumers believe cars have lower fuel costs, so firms raise prices. The perceived lower fuel costs also lead to an increase in total sales of 3.05% points. Gaming thus leads to considerable changes in the size of the market.²⁴ Consumers also substitute to more polluting cars as they substitute lower

²⁴These numbers are obtained from a model in which we assume that one out of seven households are potentially interested in purchasing a new car such that the outside good share is 61%. This results in a relatively high price elasticity of the industry. All our main results hold if we scale down the market price elasticity by changing the

Table 6: Market changes from industry wide gaming

	No Policy	Policy
	Consumer Surplus (€per household)	
Total	-24.78	208.27
Choice Distortion	-17.09	-16.72
Price Effect	-7.70	224.99
	Changes in pollution	
Total Sales (% point)	3.05	4.58
Emissions per vehicle %	0.44	5.82
	Changes in Profits	
Average Firm Profit %	8.84	15.90

Column (1) gives changes in consumer surplus, pollution and profits from 5% gaming when no policy is in place. Column (2) presents the changes when there is a binding 5%-policy in place.

fuel costs for more horsepower and size. This drives up emissions per vehicle by 0.44%. Profits increase for two reasons: firms obtain a higher markup when gaming and they have higher sales. We interpret these as short-term gains from gaming that firms may trade-off against future consumer backlash or possible legal punishment over time.

Contrast this with the column on the right in Table 6, which shows the effect of gaming when a policy is in place. Gaming now *increases* consumer surplus, with the price effect switching sign and rising in magnitude to €225. Consumers benefit from the price schedule with gaming relative to the price schedule that would result from honest compliance to the regulation. The price effect is now much higher than the choice distortion, estimated to be €17, which is similar to the corresponding value without a policy.

Note also that gaming has very large effects on pollution when there is a policy. Two forces are at play. First, the effects of the policy are undone and the goal of reducing emission is not obtained. We achieve this mechanically by letting firms game 5% in response to a 5% policy. Second, the choice distortion drives pollution per vehicle above 5%. This means that a policy with a gameable target results in more pollution relative to doing nothing. Firms benefit from gaming by avoiding compliance cost and by increasing sales and markups. The short-term profit change from gaming almost doubles from 9% to 16% when a policy is in place.

Proposition 2 in our theory showed that it was possible that consumers benefit from gaming, even when they are fooled by it. These results suggest that this is indeed the case in the European auto market. In line with Proposition 1, we find that consumers unambiguously lose from gaming when there is no policy. But, this result is overturned with the introduction of a policy. Consumers no longer benefit from honesty, which is likely to erode self regulation of the market because consumers want firms to game. Moreover, when choice distortions are significant, emissions actually *rise* in response to the policy, due to a combination of weak enforcement and significant shifts in the market due to choice distortions. We believe this to be an important finding in general, and

outside good share to 10%. We show these results in the appendix, in Table A.3.

specifically it speaks to the current state of international climate bargaining, which is based on voluntary emission saving statements rather than effective and credible measurement.

Market outcomes of asymmetric gaming: In Table 7 we compute market outcomes with asymmetry in gaming. In the left panel only one firm games, while in the right panel all but one firm games. We simulated these asymmetric cases separately for each firm, and present average impacts in the table. In all scenarios we confirm the results obtained with industry-wide symmetric gaming: gaming hurts consumers when there is no policy, while consumers benefit from gaming under the policy. Consumer surplus and pollution effects are much smaller when only one of the car makers games. Effects are almost equal to symmetric gaming when all but one firm games. The profit changes show an interesting pattern. The average gain from gaming unilaterally is a whopping 20% increase in profits, even in the absence of policy. But the profitability from gaming rises still further to 38% with policy. This shows that firms have a very strong incentive to unilaterally deviate from an honest equilibrium, and this incentive rises with policy.

The profit losses from not “gaming along” (remaining honest while all others game) are also large, which is shown in the column on the right. Being the sole honest firm leads to a 10% decrease in profits without policy and a 16% decrease with policy. Thus, if competitors start to game, it is very costly not to follow along, and this incentive is enhanced by the policy. We conclude that the policy thus clearly increases the incentive to game in a multi-firm context: the unilateral profits from deviating from honesty increase and the costs from not deviating when others do increases.²⁵

The size of our welfare effects could, of course, be sensitive to our assumptions. We discuss two important robustness checks here. First, choice distortion might increase if gaming varies across vehicles. That is, by modeling a common level of gaming on all products, we may reduce the choice distortion because most consumers will end up purchasing the same vehicle, as all vehicles will be made to have similar (false) improvements. We focus on symmetric gaming because our reduced-form evidence shows an industry-wide shift in the performance gap, but not an increase in heterogeneity. Nevertheless, it might be that gaming is more pronounced on vehicles for which it is more beneficial to game. To have an idea to what extent this determines the choice distortion, we run a simulation where gaming is drawn randomly from the uniform distribution on [0%, 10%], so that mean gaming is 5% as in the main scenario. The results are reported in the appendix, in Table A.2. We find indeed that choice distortions increase to €31, but this is still 7 times lower than the price effect of the regulation.

Second, the change in the total sales and therefore total emissions and profits depends on the price elasticity of the industry, a function of the outside good. Since we decide on the size of the outside good when fixing the potential market results might look different if we assume a larger or smaller potential market. Currently we make the assumption that one in seven European

²⁵Note that this setting is not necessarily a prisoner’s dilemma. The total industry profit from a gaming equilibrium relative to an honest equilibrium will depend on the total legal and regulatory costs of gaming relative to the total costs of compliance. The evidence presented here shows that an honest equilibrium becomes harder to support when the policy is in place.

Table 7: Market changes from asymmetric gaming

	Gaming Alone		Not Gaming Along	
	No Policy	Policy	No Policy	Policy
	Consumer Surplus (€per household)			
Total	-3.83	15.79	-24.31	189.94
Choice Distortion	-3.17	-3.24	-17.23	-16.79
Price Effect	-0.66	19.20	-7.08	206.7
	Changes in pollution			
Total Sales (% point)	0.27	0.42	2.81	4.24
Emissions per vehicle %	0.04	0.62	0.41	5.45
	Changes in Profits			
Gamer/non gamer %	20.41	37.94	-9.86	-15.83

Column (1) and (3) give changes in market equilibrium when one firm games. We present the average of letting each firm game (we run a separate simulation per firm). Changes in profits are the average for the firms that game (not for the industry). Column (2) and (4) give changes in market equilibrium when all but one firm games. We run a separate simulations for each firm not gaming along when all others game. Changes in profits are the average for the firms that do not game along.

households are on the market each year as potential car buyers which results in an outside good share of 61%. In the appendix, in Table A.3, we shrink the size of the outside good to 10%. This lowers the change in profits and sales as the overall size of the market is less elastic to changes caused from gaming. All main results are also confirmed in this setting.

6.2.1 Stated versus obtained externality savings

While the E.U. commission has released an external evaluation of the emission standard, the official ratings are still the basis for the evaluation of the program and external communications about the program's impact. In Figure 1 it is apparent that the majority of the stated decline in emissions is due to gaming. The commission claims that average emissions have fallen by 17g CO₂/km since the start of measurement in 2010.²⁶ Our estimates suggest that about 65% of the reported decline in emissions from new vehicles is due to gaming. Thus, between 2010 and 2014, the actual decrease is 5.95 grams of CO₂/km. This translates to 27 million more tons of CO₂ emissions than the naïve estimate would suggest. At \$40 per ton, this equates to \$1.11 billion per year.²⁷ Analysts who perform cost-benefit analyses of the regulation may significantly mis-evaluate the program if they calculate savings based on official ratings.²⁸

²⁶See for example <http://ec.europa.eu/clima/policies/transport/vehicles/cars>, read on 09/13/2016.

²⁷We assume that 12 million vehicles are sold across Europe, and that they are driven 14,000km per year for a 15-year period.

²⁸This is for example the case in work of one of the current co-authors, see Reynaert (2015).

7 Conclusion

Our empirical analysis demonstrates a remarkable and growing divergence between official and on-road carbon emissions for automobiles in Europe. 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 from diesel vehicles, the facts we report here suggest a veritable crisis in the administration of environmental regulations for automobiles. Moreover, combined with other findings that call into question the efficacy of other energy efficiency policies (Davis, Fuchs, and Gertler 2014, Fowle, Greenstone, and Wolfram 2015, Levinson 2016), we believe our findings point to a serious challenge for the future of climate negotiations, which are now built upon bottom-up pledges based on unreliable ex ante estimates of program conservation.

Our theoretical model offers an explanation for the evolution of emissions gaming in Europe based on Goodhart’s Law, and it gives guidance on how to assess the welfare impacts of gaming. The model points out how the introduction of corrective policies can alter the allegiance of consumers, who dislike gaming in the absence of policy, but may ultimately benefit (privately) from gaming once a policy is introduced. In addition, our model points out how policy-induced gaming alters the efficiency of private transactions by inducing choice distortion. Our structural estimation verifies the economic importance of these effects and shows that price effects operating through cost pass-through dominate choice distortion, which implies that consumers do indeed privately benefit from gaming. It is important to emphasize, however, that this private benefit will be outweighed by the failure to mitigate the externality, so long as the policy was well designed to begin with.

We anticipate that these same concerns are present in a variety of regulatory settings, and we expect that both our theoretical and empirical approach can be applied elsewhere. Critical to our setup is that the measure which can be gamed is used by the regulator *and* is instrumental to consumer choice. For example, in education, test scores of current students might be gamed to satisfy policy mandates that affect resource transfers. The test scores also affect the demand for admission from future waves of students. Consumers (students) can benefit from gaming when gaming frees up resources at a school that would otherwise be expended to satisfy policy and obtain transfers. Or, in finance, capital requirements can be used to minimize an individual bank’s role in creating systemic risk. Customers can privately benefit if a bank is able to rebalance its portfolio more profitably after gaming the regulated measures.

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>.
- Allcott, Hunt and Nathan Wozny. 2014. “Gasoline Prices, Fuel Economy, and the Energy Paradox.” *Review of Economics and Statistics* 96 (5):779–795.

- Anderson, Soren T. and James M. Sallee. Forthcoming. “Designing Policies to Make Cars Greener: A Review of the Literature.” *Annual Review of Resource Economics* .
- Barnett, A.H. 1980. “The Pigouvian Tax Rule Under Monopoly.” *American Economic Review* 70 (5):1037–1041.
- Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. “Automobile Prices in Market Equilibrium.” *Econometrica* 64 (4):841–890.
- Booz & Company. 2012. “Company Cars and Fring Benefit Tax: Understanding the Impacts on Strategic Transport Targets.” NZ Transport Agency research report 474.
- Buchanan, James M. 1969. “External Diseconomies, Corrective Taxes, and Market Structure.” *American Economic Review* 59 (1):174–177.
- Busse, Meghan R., Christopher R. Knittel, and Florian Zettelmeyer. 2013. “Are Consumers Myopic? Evidence from New and Used Car Purchases.” *American Economic Review* 103 (1):220–256.
- 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.
- Davis, Lucas W., Alan Fuchs, and Paul Gertler. 2014. “Cash for Coolers: Evaluating a Large-Scale Appliance Replacement Program in Mexico.” *American Economic Journal: Economic Policy* 6 (4):207–238.
- 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.
- Gabaix, Xavier and David Laibson. 2006. “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets.” *Quarterly Journal of Economics* 121 (2):505–540.
- Gerlagh, Reyer, Inge van den Bijgaart, Hans Nijland, and Thomas Michielsen. Forthcoming. “Fiscal policy and CO2 emissions of new passenger cars in the EU.” *Environmental and Resource Economics* .
- 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=ZGVmYXVsdGRvbWFpbnxmcmFua3ZlcmJvfgd40jZiMzQ4YzUwYjhiNjdjOTk>. Manuscript: University of Leuven.
- Heidhues, Paul, Botond Köszegi, and Takeshi Murooka. 2016. "Inferior Products and Profitable Deception." *Review of Economic Studies* .
- 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.
- Langer, Ashley and Shaun McRae. 2014. "Step on It: Evidence on the Variation in On-Road Fuel Economy." Manuscript: University of Arizona.
- Leggett, Christopher G. 2002. "Environmental Valuation with Imperfect Information." *Environmental and Resource Economics* 23:343–355.
- Levinson, Arik. 2016. "How Much Energy Do Building Energy Codes Save? Evidence from California Houses." *American Economic Review* 106 (10):2867–2894.
- 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.
- Reynaert, Mathias and Frank Verboven. 2014. "Improving the performance of random coefficients demand models: The role of optimal instruments." *Journal of Econometrics* 179 (1):83–98. URL <https://ideas.repec.org/a/eee/econom/v179y2014i1p83-98.html>.
- Sallee, James M. 2014. "Rational Inattention and Energy Efficiency." *Journal of Law and Economics* 57 (3):781–820.

- Sallee, James M., Sarah E. West, and Wei Fan. 2016. "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations." *Journal of Public Economics* 135:61–73.
- 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 CO₂ 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 CO₂ 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.

A Appendix: Additional Tables

Table A.1: 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 (L/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.

Table A.2: Market Changes from Gaming - Heterogenous Gaming

	All Game	Own Gaming	All other game
	No Policy		
	Consumer Surplus (€per household)		
Total	-39.74	-5.21	-38.14
Choice Dist.	-31.10	-4.46	-30.18
Price Effect	-8.64	-0.75	-7.95
	Emissions % changes		
Sales	3.06	0.26	2.82
Per vehicle	0.53	0.05	0.49
	Profits % changes		
Profit	8.68	20.37	-9.77
	Policy		
	Consumer Surplus (€per household)		
Total	186.29	13.82	169.54
Choice Dist.	-30.74	-4.63	-29.83
Price Effect	217.04	18.54	199.37
	Emissions % changes		
Sales	4.52	0.42	4.19
Per vehicle	6.19	0.65	5.78
	Profits % changes		
Profit	15.29	37.35	-15.84

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

Table A.3: Market Changes from Gaming - Low Outside Good Share

	All Game	Own Gaming	All other game
No Policy			
Consumer Surplus (€per household)			
Total	-6.61	-8.32	-12.69
Choice Dist.	-8.73	-8.77	-14.88
Price Effect	2.12	0.45	2.19
Emissions % changes			
Sales	0.49	0.04	0.45
Per vehicle	0.56	0.05	0.52
Profits % changes			
Profit	1.25	21.43	-16.79
Policy			
Consumer Surplus (€per household)			
Total	627.60	46.83	572.11
Choice Dist.	-8.60	-9.18	-14.00
Price Effect	636.20	56.01	586.12
Emissions % changes			
Sales	0.51	0.05	0.47
Per vehicle	5.68	0.60	5.31
Profits % changes			
Profit	2.63	38.46	-26.01

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

B Appendix: Proofs

Proposition 1. *In the absence of policy ($\sigma = \infty, \tau = 0$), consumer surplus falls with the level of gaming. Specifically:*

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

We derive this by starting with the definition of consumer surplus as the integral under the inverse demand curve. Because the true attribute of the good x is unaffected by gaming, we can analyze consumer surplus using the true demand curve. The standard portion of consumer surplus is the integral from the final price, denoted p^* up to infinity. Denote by \tilde{p} the upfront purchase price that would induce a sophisticated consumer to purchase the amount of the good that is in fact purchased at price p^* by the consumer with perception $(1-\alpha)$. The choice distortion can be written as the difference between the revenue generated between p^* and \tilde{p} and the consumer value generated between those points.

$$CS = \underbrace{\int_{p^*}^{\infty} D(z + \beta x) dz}_{\text{CS of correct quantity}} + \underbrace{\int_{\tilde{p}}^{p^*} D(z + \beta x) dz}_{\text{Value of excess quantity}} - \underbrace{\int_{\tilde{p}}^{p^*} D(\tilde{p} + \beta x) dz}_{\text{Cost of excess quantity}}. \quad (\text{B.1})$$

Differentiation of equation B.5, in which p^* and \tilde{p} are endogenously determined by g , yields the result. Note that the inside of the third integral is a constant with respect to the variable of integration, so it can be pulled out of the integral, leaving only the constant 1 inside. Specifically:

$$\begin{aligned} \frac{dCS}{dg} = & -D(p^* + \beta x) \frac{dp^*}{dg} + \left\{ D(p^* + \beta x) \frac{dp^*}{dg} - D(\tilde{p} + \beta x) \frac{d\tilde{p}}{dg} \right\} \\ & - \left\{ (p^* - \tilde{p}) D'(\tilde{p} + \beta x) \frac{d\tilde{p}}{dg} + D(\tilde{p} + \beta x) \left(\frac{dp^*}{dg} - \frac{d\tilde{p}}{dg} \right) \right\}. \end{aligned} \quad (\text{B.2})$$

Using the pass through coefficient ρ , a change in g scales to a change in tax by $(1-\alpha)\beta$, so $dp^*/dg = \rho(1-\alpha)\beta$. This simplifies the first term to yield the result.

For the second term in B.2, note that $\tilde{p} = p^* - (1-\alpha)\beta g$ by definition. Then, substitute a first-order Taylor approximation to write demand at \tilde{p} as a function of demand at p^* and $D' \times (\tilde{p} - p^*)$:

$$\begin{aligned} & D(p^* + \beta x) \frac{dp^*}{dg} - D(\tilde{p} + \beta x) \frac{d\tilde{p}}{dg} \\ = & D(p^* + \beta x) \frac{dp^*}{dg} - D(p^* + \beta x - (1-\alpha)\beta g) \frac{d\tilde{p}}{dg} \\ \approx & \left\{ D(p^* + \beta x - (1-\alpha)\beta g) + D'(p^* + \beta x - (1-\alpha)\beta g)(1-\alpha)\beta \right\} \frac{dp^*}{dg} - D(p^* + \beta x - (1-\alpha)\beta g) \frac{d\tilde{p}}{dg} \\ = & D(p^* + \beta x - (1-\alpha)\beta g) \left(\frac{dp^*}{dg} - \frac{d\tilde{p}}{dg} \right) + D'(p^* + \beta x - (1-\alpha)\beta g)(1-\alpha)\beta \frac{dp^*}{dg} \end{aligned} \quad (\text{B.3})$$

Now consider the third term in B.2. After substituting $\tilde{p} = p^* - (1 - \alpha)\beta g$, we see that the term that multiplies the difference in derivatives will cancel in the third term of B.2 and the second-term, defined using B.3. This means that B.2 can be written:

$$\begin{aligned} \frac{dCS}{dg} &= -D(p^* + \beta x)\rho(1 - \alpha)\beta + D'(p^* + \beta x - (1 - \alpha)\beta g)(1 - \alpha)\beta \frac{dp^*}{dg} \\ &\quad - \left\{ (p^* - \tilde{p})D'(\tilde{p} + \beta x) \frac{d\tilde{p}}{dg} \right\}. \end{aligned} \quad (\text{B.4})$$

Substitute $p^* - \tilde{p} = (1 - \alpha)\beta g$ and $d\tilde{p}/dg = (\rho - 1)(1 - \alpha)\beta$, which follows from differentiating the definition of \tilde{p} and using the pass through result for p^* . Then, simplification yields the final result:

$$\begin{aligned} \frac{dCS}{dg} &= -D(p^* + \beta x)\rho(1 - \alpha)\beta + D'(p^* + \beta x - (1 - \alpha)\beta g)(1 - \alpha)\beta \rho(1 - \alpha)\beta \\ &\quad - \left\{ (1 - \alpha)\beta g D'(\tilde{p} + \beta x)(\rho - 1)(1 - \alpha)\beta \right\} \\ &= -D(p^* + \beta x)\rho(1 - \alpha)\beta + D'(p^* + \beta x - (1 - \alpha)\beta g)(1 - \alpha)^2 \beta^2 \rho \\ &\quad - (1 - \alpha)^2 \beta^2 g D'(\tilde{p} + \beta x)(\rho - 1) \\ &= -D(p^* + \beta x)\rho(1 - \alpha)\beta + (1 - \alpha)^2 \beta^2 g D'(\tilde{p} + \beta x). \end{aligned}$$

Note that, with the local linear demand assumption, the derivative of D evaluated at either \tilde{p} or p^* is the same. ■

Proposition 2. *In the presence of a binding standard but no tax ($\lambda > 0, \tau = 0$), a change in gaming affects consumer surplus as follows:*

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

Proposition 2 is derived in a similar way to Proposition 1, but we define surplus using integrals over the demand function starting with full prices f . (This same could have been done in the prior proof, yielding the same result.) Recall that $f = p + \beta x$ and $\tilde{f} = p + \beta x - (1 - \alpha)\beta g$.

$$CS = \underbrace{\int_{f^*}^{\infty} D(z)dz}_{\text{CS of correct quantity}} + \underbrace{\int_{\tilde{f}}^{f^*} D(z)dz}_{\text{Value of excess quantity}} - \underbrace{\int_{\tilde{f}}^{f^*} D(\tilde{f})dz}_{\text{Cost of excess quantity}}. \quad (\text{B.5})$$

Differentiating, using the same Taylor approximation as above to simplify, yields:

$$\frac{dCS}{dg} \approx -D(f^*) \frac{df^*}{dg} + D'(f^*)(f^* - \tilde{f}) \left(\frac{df^*}{dg} - \frac{d\tilde{f}}{dg} \right).$$

Substitute the pass through result described in the text: $df^*/dg = \rho(c'(x) + \alpha\beta) + \beta$. Substitute

the definition of \tilde{f} , which shows that $f^* - \tilde{f} = (1 - \alpha)\beta g$. And substitute $\frac{df^*}{dg} - \frac{d\tilde{f}}{dg} = (1 - \alpha)\beta$:

$$\begin{aligned}\frac{dCS}{dg} &\approx -D(f^*)(\rho(c'(x) + \alpha\beta) + \beta) + D'(f^*) \{(1 - \alpha)\beta g\} \{(1 - \alpha)\beta\} \\ &= -D(f^*)(\rho(c'(x) + \alpha\beta) + \beta) + D'(f^*)(1 - \alpha)^2\beta^2 g.\end{aligned}\tag{B.6}$$

This yields the result. ■

Proposition 3. *In the presence of a tax but no standard ($\sigma = \infty, \tau > 0$), consumer surplus falls with the level of gaming. Specifically:*

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

Proposition 4 follows directly from Proposition 1 once it is recognized that this is the exact same problem with β replaced by $\beta + \tau$ at all points. Thus, the exact same proof as Proposition 1 holds, yielding the result. ■

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 $\sigma > \sigma^*$.*

With g fixed, $x = \sigma - g$. And, pass through will be

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 $\lambda > 0$.*

This corollary follows immediately from the result in Proposition 2 that defines dCS/dg , noting that the choice distortion term is 0 when $g = 0$. The setup of the corollary says that the price effect is negative, which delivers the result immediately. ■

Proposition 4. *Goodhart's Law for externality-correcting policies: a tighter standard induces greater gaming:*

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