OUTPUT AND ATTRIBUTE-BASED CARBON REGULATION UNDER UNCERTAINTY

Ryan Kellogg

Output and Attribute-Based Carbon Regulation Under Uncertainty
Ryan Kellogg
NBER Working Paper No. 26172
August 2019
JEL No. D81,Q54,Q58

## ABSTRACT

Output-based carbon regulations—such as fuel economy standards and the rate-based standards in the Clean Power Plan—create well-known incentives to inefficiently increase output. Similar distortions are created by attribute-based regulations. This paper demonstrates that, despite these distortions, output and attribute-based standards can always yield greater expected welfare than "flat" emission standards given uncertainty in demand for output (or attributes), assuming locally constant marginal damages. For fuel economy standards, the welfare-maximizing amount of attribute or mileage-basing is likely small relative to current policy. For the electricity sector, however, an intensity standard may yield greater expected welfare than a flat standard.

Ryan Kellogg
University of Chicago
Harris School of Public Policy
1307 East 60th Street
Chicago, IL 60637
and NBER
kelloggr@uchicago.edu

# 1  Introduction

Environmental economists have long extolled the virtues of Pigouvian taxation or cap-and-trade pollution permit systems as efficient public policies for correcting environmental externalities, and especially for addressing carbon emissions. While such programs have begun to appear—the European Union's and California's cap-and-trade programs and British Columbia's carbon tax are notable examples—many policies in practice deviate from economists' favorite prescription. In the United States, carbon regulations frequently express the maximum permissible quantity of emissions (or minimum permissible quantity of a "clean" input) as a function of the output of an underlying good or of attributes such as the good's size or weight. For instance, state-level renewable portfolio standards mandate that renewable electricity generation meet a minimum share of total generation (or capacity), fuel economy standards mandate a maximum amount of fuel use in proportion to vehicle miles traveled, and the Obama Clean Power Plan allowed states to regulate $CO_2$ emissions as a proportion of total electricity generation.

Output-based standards generate well-understood incentives to distort output, as first recognized in Kwoka (1983). Holland, Hughes, and Knittel (2009) makes this point clearly in the context of California's Low Carbon Fuel Standard, an intensity standard on the embodied carbon in transportation fuel that creates an incentive to produce more fuel overall, since doing so slackens the constraint on $CO_2$ emissions. This intuition applies broadly. Fuel economy standards, for example, lead to the well-known "rebound effect" that causes an increase in vehicle miles traveled (see Gillingham (forthcoming) for a survey of the recent literature on rebound). As for settings where the emissions standard is a function of goods' attributes rather than output, Ito and Sallee (2018) shows that basing fuel economy standards on vehicles' weight or footprint results in inefficiently over-sized vehicles.

In this paper, I ask whether output or attribute-based emissions regulation can actually be economically advantageous relative to a "flat" emissions standard in the presence of uncertainty about future abatement costs. The intuition is straightforward. Under cap-and-

trade, the quantity of emissions is fixed so that shocks, whether to demand or supply for the underlying good, or to abatement cost itself, cause fluctuations in marginal abatement cost. Assuming that the marginal damage from carbon emissions is locally constant—a reasonable assumption for a global stock pollutant like $CO_2$ (Newell and Pizer 2003)—this variation reduces welfare relative to the first-best, following the arguments from Weitzman (1974). Output or attribute-based regulations can dampen these marginal abatement cost fluctuations and their associated welfare losses because emissions are allowed to vary as shocks occur.[1]

Can the flexibility benefits from output or attribute-based standards outweigh the welfare losses from the distortions that they impose? To the best of my knowledge, this question has not been addressed despite the ubiquity of these policies. Holland, Hughes, and Knittel (2009) briefly notes that intensity standards may yield welfare benefits in an uncertain environment, but because the model in that paper is deterministic it does not explore this point in depth. Similarly, Anderson and Sallee (2016) and Ito and Sallee (2018) raise but do not answer the question of whether attribute-based fuel economy standards are advantageous relative to a non-attribute-based standard in the presence of uncertainty.

I begin by building a partial equilibrium model in which there is a single good, the production or consumption of which is associated with $CO_2$ emissions. $CO_2$ emissions have a constant external marginal cost as well as private costs and benefits (e.g., increases in emissions may be associated with larger fuel costs and lower expenditures on energy-efficient capital). Consumers and producers in equilibrium choose both the emissions level and either the good's quantity or the value of one of its attributes. The model can therefore address emissions standards that are a function of output (like the Low Carbon Fuel Standard) and standards that are a function of an attribute (like footprint-based U.S. fuel economy standards). Thus, output and attribute-based standards are in fact conceptually equivalent, a point that has gone unnoticed in prior work.

---

[1]Compliance flexibility was, in fact, one of the justifications offered for the U.S. footprint-based fuel economy standards. See, for instance, Lutsey (2015).

Agents' demand for output and emissions in the model is affected by two exogenous shocks that are uncertain when the regulator sets its policy. First, there may be direct shocks to marginal abatement cost—agents' willingness to pay for emissions reductions—holding output (or attributes) fixed. Fuel prices are perhaps the most salient example of this form of uncertainty. Second, there may be shocks to the marginal value of the good's output or attribute, holding emissions fixed. For instance, income shocks may lead agents to want to consume more electricity, more miles traveled, or larger vehicles. These shocks lead indirectly to uncertainty about marginal abatement cost, since marginal abatement cost typically increases with output, holding emissions fixed.

Under a flat emissions standard, either form of uncertainty leads to an expected welfare loss relative to the first-best, since marginal abatement cost diverges from marginal damage. The first-best welfare outcome can be achieved by a Pigouvian tax or by an emissions standard that is indexed directly to the exogenous source of uncertainty (gasoline prices, for example, as shown in Kellogg (2018)).[2] But what happens to expected welfare if the emissions standard is instead a function of the endogenously-determined output or attribute of the good, as is common in current policy?

I show that the welfare properties of output or attribute-based standards depend crucially on which type of uncertainty is present. If uncertainty is due only to shocks to marginal abatement cost, basing an emissions standard to the good's output or attributes always strictly reduces welfare relative to a flat standard. This result is a generalization of the finding in Kellogg (2018) that an attribute-based fuel economy standard strictly reduces welfare when the only source of uncertainty is the future price of gasoline. The intuition for this result stems from the fact that both the welfare cost of the distortion to the good's output or attribute and the welfare benefit from increased flexibility are second-order. I show that the former effect dominates.

---

[2]Pizer and Prest (2016) shows that the first-best outcome can also potentially be achieved using a non-indexed emissions standard with retroactive quantity updating and banking and borrowing provisions that extend beyond the policy revision horizon.

When there are shocks to the marginal value of output, however, expected welfare can always be strictly increased by at least a small amount of output-basing (the same argument holds for attributes). This result is driven by the fact that the flexibility benefits of an output-based standard are now first-order, while the welfare costs of the induced distortions remain second-order. I show that the optimal amount of output-basing increases with: (1) uncertainty in future output (relative to the magnitude of the unpriced externality); (2) the extent to which emissions are affected by output shocks in the absence of regulation; and (3) the inelasticity of output with respect to the good's cost. It is, moreover, possible that an intensity standard—an emissions standard that is a linear rather than merely affine function of output—yields greater expected welfare than a flat standard.

Is this possibility result important in practice? I quantitatively evaluate the welfare effects of output and attribute-standards by calibrating the model to three policy settings: (1) an attribute-based fuel economy standard (holding miles traveled fixed); (2) a fuel economy standard with endogenous miles traveled (holding vehicle attributes fixed); and (3) an output-based $CO_2$ standard for electric generation. In each case, I calibrate the model using estimates from previous studies, and I estimate uncertainty in the net demand for output and attributes using their historic volatility.

I find that, even with generous parameter assumptions and estimates of uncertainty in the demand for vehicle size, the U.S. footprint-based fuel economy standard yields lower expected welfare than a "flat" non-attribute based standard, and that the welfare-maximizing footprint-based standard is nearly flat. Similarly, when I treat miles driven as endogenous, I find that a regulation imposing a ceiling on total gasoline consumption yields greater expected welfare than a fuel economy standard. These results are mostly driven by the fact that uncertainty about the future demand for miles traveled, and especially about the demand for vehicle size, is fairly small.

In contrast, I find that the welfare-maximizing emissions standard for the U.S. electricity market is substantially output-based. Depending on parameter inputs, an intensity standard

may even deliver greater expected welfare than a flat standard. This result is driven by the large uncertainty about the future demand for electric power and by the fact that emissions vary substantially with output shocks. These industry features are also documented in Borenstein *et al.* (forthcoming), which finds substantial uncertainty in the future business-as-usual emissions path for California's electricity sector.

My result that output and attribute-based standards can increase expected welfare relates to other work that has raised normative justifications for intensity standards. Holland (2012) and Fowlie, Reguant, and Ryan (2016), for instance, show that in the presence of market power or unregulated emissions in substitute sectors, intensity standards can welfare-dominate a flat standard (or a Pigouvian tax), since the implied subsidy to output mitigates the pre-existing market distortions. This paper can be viewed as an extension of this second-best logic to settings where flat standards (though not Pigouvian taxes) fail to achieve the first-best due to uncertainty in the economic environment.

This paper also relates to previous research that has studied intensity *targets*. Ellerman and Wing (2003), Quirion (2005), Newell and Pizer (2008), Heutel (2012), and Zhao (2018) consider the consequences of indexing aggregate $CO_2$ emission limits to GDP and conclude that indexing can improve expected welfare relative to a non-indexed policy.[3] This literature, unlike the modeling framework I develop here, models intensity targets such that they do not generate output distortions. Zhao (2018), for instance, models a policy that adjusts an aggregate emissions cap each year in response to GDP shocks but is implemented as a standard cap-and-trade program, so that firms do not have an incentive to distort their output. Such a policy would dominate the distortion-inducing regulations that I model here. In practice, however, emissions standards are overwhelmingly indexed to quantities and attributes that are manipulable by firms or consumers rather than to exogenous variables such as GDP or fuel prices. Renewable portfolio standards, energy efficiency standards, fuel

---

[3]Indexed regulation is not always superior to non-indexed regulation in these papers because they model either an upward-sloping marginal damage function (appropriate for local pollutants or for global, long-run $CO_2$ regulation) or noise in GDP measurements that is not associated with emissions.

economy standards, the Clean Power Plan, and California's Low Carbon Fuel Standard all generate incentives to distort output or goods' attributes.

The paper proceeds as follows. Section 2 introduces the modeling framework, and then section 3 delivers the main theoretical results of the paper: output or attribute-based emissions standards never increase expected welfare when only the marginal abatement cost is uncertain, but they do improve welfare relative to a "flat" standard when the marginal value of output (or an attribute) is uncertain. Section 4 examines the quantitative implications of this result in three policy applications. Section 5 concludes.

# 2   Model setup

This section introduces the model that I use throughout the paper. The model is similar to that in my previous work on fuel economy and gasoline price uncertainty (Kellogg 2018). Because the main objective of the model is to distill intuition, I abstract away from other potential market failures such as market power, emissions leakage, or incomplete information.

The model considers a single carbon-emitting good that is supplied by competitive firms and demanded by price-taking, fully-informed consumers. In what follows, I frequently refer to firms and consumers collectively as "agents". The exposition below is framed in terms of agents' choices over the good's quantity, but with a simple re-labeling (replace every "$Q$" with an "$A$"), the model could instead apply to choices over one of the good's attributes.

## 2.1   Private agents' supply, demand, and equilibrium

The good's overall output (and consumption) is given by $Q$, and $CO_2$ emissions are denoted by $E$. Agents' choices of $Q$ and $E$ are affected by two additional variables: $\eta$ and $F$. $\eta$ denotes factors that shift preferences for (or costs of) additional output $Q$, and $F$ denotes factors that affect the private costs of additional emissions $E$.

The model is designed to be as general as possible regarding the specific objects rep-

resented by $Q$, $E$, $\eta$, and $F$. But to briefly fix ideas, the three applications considered in section 4 will use them to represent:

1. **Attribute-based fuel economy standards:** $Q$ denotes the average footprint of vehicles sold, $E$ denotes the average fuel economy of vehicles sold, $\eta$ represents taste or income shocks that affect the marginal value of footprint, and $F$ denotes the price of fuel. Miles traveled are fixed.

2. **Fuel economy standards with endogenous miles traveled:** $Q$ denotes average lifetime miles traveled per vehicle, $E$ denotes lifetime average gasoline use per vehicle, $\eta$ represents taste or income shocks that affect the marginal value of driving, and $F$ denotes the price of fuel. Vehicle attributes are fixed.

3. **Electricity generation:** $Q$ denotes total electricity consumption, $E$ denotes total $CO_2$ emissions, $\eta$ represents taste or income shocks that affect the marginal value of electricity, and $F$ denotes the relative price of high-emission versus low-emission generation fuels.

Similar to Kellogg (2018), I posit a private net benefits function $B(Q, E, \eta, F)$ that captures the difference between consumers' utility (not including any externalities) and firms' production cost. I assume that $B(Q, E, \eta, F)$ is twice continuously differentiable and, in the spirit of Weitzman (1974), that it can be well-approximated by a second-order Taylor expansion. This assumption substantially enhances the analytic tractability of the model, and I maintain it throughout the paper. I show in appendix B that $B(Q, E, \eta, F)$ can be micro-founded as a summation of individual agents' private benefit functions, and that it will be a sufficient statistic for utilitarian total private welfare under the regulations I consider given: (1) restrictions on heterogeneity in agents' valuations of $Q$ and $E$; (2) inclusion of compliance credit trading in any emissions standard; and (3) equal and constant marginal utility of income (and welfare weights) across all agents.[4]

---

[4]The last of these restrictions rules out distributional preferences as a rationale for output or attribute-based policies.

I assume that there are diminishing net returns to both $Q$ and $E$ so that, letting $B_{QQ}$ and $B_{EE}$ represent the second derivatives of the net benefit function, $B_{QQ} < 0$ and $B_{EE} < 0$. For example, consumers receive declining marginal utility from miles driven, and reducing vehicles' fuel use per mile becomes increasingly costly as vehicles become more and more fuel efficient. I also assume that the cross-partial $B_{QE} \geq 0$ to reflect the fact that in most applications, reducing emissions is more costly when $Q$ is large.

I define $\eta$ and $F$ so that $B_{Q\eta} > 0$, $B_{QF} = 0$, $B_{E\eta} = 0$, and $B_{EF} < 0$. $\eta$ should be thought of as an income or preference shock that affects demand for $Q$, though it could also represent a marginal production cost shifter. $F$ should be thought of as the price ($=$ marginal cost) of carbon-intensive fuel, where increased fuel prices decrease the net marginal benefit from emissions. It could also, in principle, represent the state of firms' emissions control technology. The restriction that $F$ does not directly affect the demand for $Q$ is consistent with an assumption that consumers have quasilinear utility in fuel expenditures.[5]

Absent regulation, agents in equilibrium will choose $Q$ and $E$ so that the private first-order conditions (FOCs) are satisfied: $U_Q(Q, E, \eta, F) = 0$ and $U_E(Q, E, \eta, F) = 0$. In addition to the assumptions above, sufficient conditions for an interior equilibrium are: (1) $B_Q(0, E, \eta, F) > 0$ and $B_E(Q, 0, \eta, F) > 0$ (i.e., consumers' marginal utilities of $Q$ and $E$ at zero are strictly greater than firms' marginal costs); and (2) satisfaction of the second-order condition (SOC) $B_{QQ}B_{EE} - B_{QE}^2 > 0$. In addition, application of the implicit function theorem yields the comparative statics $dQ/d\eta > 0$, $dQ/dF \leq 0$, $dE/d\eta \geq 0$, and $dE/dF < 0$, with all inequalities strict iff $B_{QE} > 0$.[6]

---

[5]The quasilinearity assumption is common in empirical vehicle fuel economy choice models, such as Busse, Knittel, and Zettelmeyer (2013), Allcott and Wozny (2014), and Sallee, West, and Fan (2016).

[6]Formulas for these derivatives are derived and presented in appendix A.

## 2.2 The externality and regulatory responses

The total external cost of emissions is given by $\phi E$, so that marginal damage is equal to $\phi$, a constant that is proportional to the social cost of carbon.[7] The full-information social planner's problem is then given by:

$$\max_{Q,E} B(Q, E, \eta, F) - \phi E \tag{1}$$

The planner's FOCs are given by $B_Q(Q, E, \eta, F) = 0$ and $B_E(Q, E, \eta, F) = \phi$,[8] and the SOC is identical to that in the private agents' problem. Intuitively, the planner's solution involves a lower $E$ than the private optimum, and then conditional on the choice of $E$ the planner chooses the same $Q$ as private agents.

I model the regulator's problem in two stages. In the first stage, the regulator must commit to a policy without knowing what the realized values of $\eta$ and $F$ will be in the second stage (the parameters governing the benefits function are, however, common knowledge). $V(\eta)$ and $W(F)$ denote the regulator's stage 1 rational beliefs about the distributions of $\eta$ and $F$, which have support on $[\eta_L, \eta_H]$ and $[F_L, F_H]$, respectively. I denote the density functions by $v(\eta)$ and $w(F)$, and I denote the expected values of $\eta$ and $F$ by $\bar{\eta}$ and $\bar{F}$. Then in stage 2 agents choose $Q$ and $E$ given the policy and the realized $\eta$ and $F$. I model stage 2 as a single compliance period, though it would be straightforward to the extend the model to multiple compliance periods, each with a different realization of $\eta$ and $F$.

---

[7]Per Weitzman (1974), and as in Kellogg (2018), the results of the paper are unchanged if $\phi$ is defined as the expected value of uncertain marginal damages, so long as the stochastic component of marginal damage is uncorrelated with $\eta$ or $F$. In this case, what the paper describes as "first-best" is first-best subject to incomplete information about marginal damage.

[8]I assume that $B_E(Q, 0, \eta, F) > \phi$, so that the planner's problem has an interior solution.

# 3  When do output and attribute-based standards improve expected welfare?

It is easy to show that a Pigouvian tax $\tau$ on emissions $E$ achieves the full-information first-best, since in that case the agents' welfare maximization problem is identical to that of the social planner. What if a tax policy is not available, and the regulator must instead use an emissions standard? This section considers the welfare effects of an output-based standard, in which emissions $E$ are constrained to lie below a function $\mu(Q)$ of output. I focus attention on affine output-based standards $\mu(Q) = \mu_0 + \gamma Q$, which encompass "flat" standards whenever $\gamma = 0$ and intensity standards whenever $\mu_0 = 0$ and $\gamma > 0$.

What are the welfare-maximizing levels of $\mu_0$ and $\gamma$? In particular, is the optimal regulation output-based, so that $\gamma > 0$? I first show that if there is only uncertainty over $F$, then optimal regulation requires $\gamma = 0$. The argument follows that in Kellogg (2018) for why attribute-based fuel economy standards are never optimal in the face of gasoline price uncertainty. The heart of this section of the paper then considers regulation when there is uncertainty over $\eta$. I show that the welfare-maximizing value of $\gamma$ is generically non-zero and will be strictly greater than zero unless $\eta$ and $F$ have a large positive correlation.

## 3.1  Optimal standards under uncertainty in $F$

Suppose that $\eta$ is fixed and known by the regulator in period 1, so that there is no uncertainty in the marginal value of $Q$ (I henceforth suppress $\eta$ in the notation for the remainder of this subsection). What are the welfare-maximizing regulatory parameters $\mu_0$ and $\gamma$?[9] It is useful to begin by building intuition for agents' choices of $Q$ and $E$ in the absence of regulation. Appendix A provides formal comparative statics for how $Q$ and $E$ vary with $F$, and these relationships are illustrated by the upward-sloping line in $Q$-$E$ space in figure 1, panel (a).

---

[9]This subsection closely follows section 6 in Kellogg (2018), replacing $a$ from that paper with $Q$, and replacing $G$ with $F$.

Intuitively, both output and emissions decrease as $F$ rises from $F_L$ to $F_H$.[10]

At the social optimum, agents internalize the externality $\phi$. The effects of doing so on $Q$ and $E$ are equivalent to an increase in $F$, so the socially optimal $Q$ and $E$, for a given value of $F$, are simply given by a southwestward shift away from the agents' preferred $Q$ and $E$ along the choice line in figure 1, panel (a).

Figure 1, panel (b) imposes a "flat" non-output-based standard that is optimally set at $\mu_0 = \mu^*$ and binds whenever $F \leq \hat{F}(\mu^*)$. Following Kellogg (2018), the optimality of $\mu^*$ implies that it satisfies the FOC that its marginal compliance cost, conditional on the standard binding, equals marginal external damage. Formally, $\mu^*$ satisfies:

$$\frac{1}{W(\hat{F}(\mu^*))} \int_{F_L}^{\hat{F}(\mu^*)} B_E(Q(\mu^*, F), \mu^*, F) w(F) dF = \phi. \tag{2}$$

Note that when the flat $\mu^*$ standard binds, agents' choice of $Q$ is always given by $Q^*$, as shown in figure 1, panel (b), regardless of the realization of $F$.[11] Equivalently, the level set of agents' private welfare $B(Q, E, F)$ that corresponds to agents' optimal choice of $Q$ and $E$ when constrained by the standard always has a point of tangency at $(Q^*, \mu^*)$ for any $F < \hat{F}$.

Panels (c) and (d) of figure 1 illustrate agents' choices under an output-based standard that rotates the flat $\mu^*$ standard around the point $(Q^*, \mu^*)$. When $\gamma > 0$, agents' choices of $Q$ and $E$ now vary with $F$ when the standard binds, since increasing $Q$ allows agents to increase $E$. The covariance between $Q$ and $F$ is proportional to $\gamma$ (see appendix A for formal comparative statics), so that little variation in $Q$ and $E$ is induced when $\gamma$ is small (panel (c) of figure 1), but variation in $Q$ and $E$ is substantial when $\gamma$ is large (panel (d)).

---

[10]$Q$ and $E$ are affine in $F$ under the second-order Taylor expansion assumption from section 2.1.

[11]This result follows from the construction that $B_{QF} = 0$.

**Figure 1:** Agents' choices, social optima, and regulations under uncertainty in $F$

**(a)** Agents' unrestricted choices and social optima
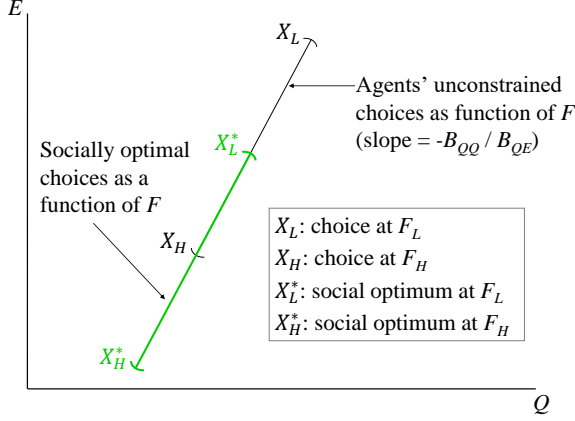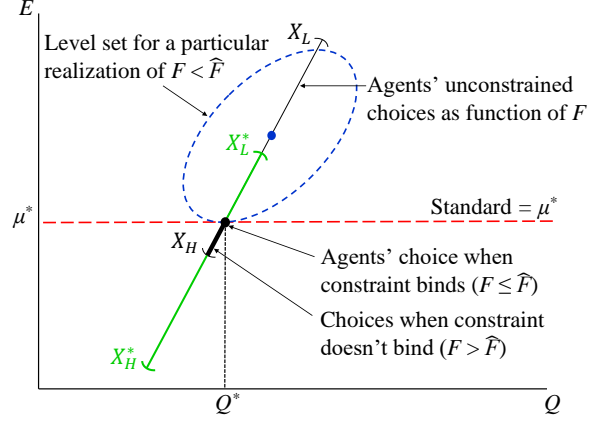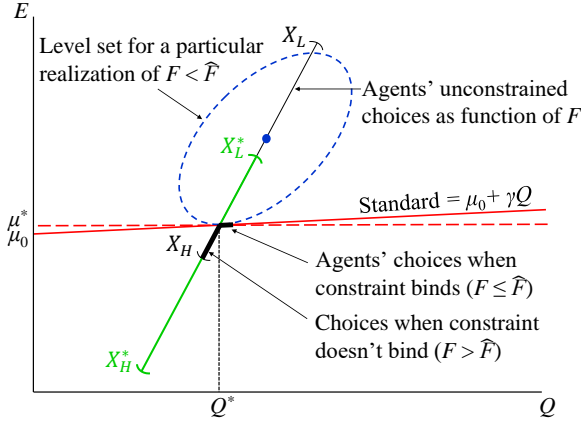


**(b)** Optimal flat standard



**(c)** Slightly output-based regulation



**(d)** Heavily output-based regulation
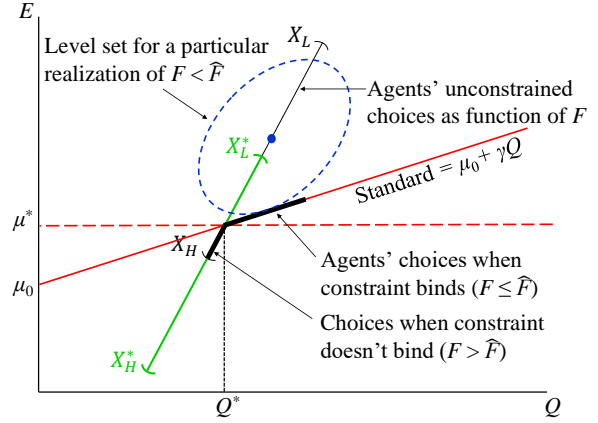


Note: $Q$ denotes the good's output (or an attribute), and $E$ denotes emissions. When unconstrained by regulation, agents' choices lie on the thin upward-sloping black line between the points $X_L$ and $X_H$, where a lower realized fuel price $F$ leads to higher choices of $Q$ and $E$. The socially optimal set of outcomes is given by a southwestward shift along this line, so that socially optimal choices lie between $X_L^*$ and $X_H^*$. $F_L$ and $F_H$ denote the lowest and highest possible fuel prices, respectively. The output-based standard $\mu(Q) = \mu_0 + \gamma Q$ binds only for fuel prices less than $\hat{F}(\mu_0, \gamma)$. See text for details.

The optimal values for $\mu_0$ and $\gamma$ come from solving the regulator's problem:

$$\max_{\mu_0, \gamma} \int_{F_L}^{\hat{F}(\mu_0,\gamma)} \left( B(Q(\mu_0, \gamma, F), E(\mu_0, \gamma, F), F) - \phi E(\mu_0, \gamma, F) \right) w(F) dF$$

$$+ \int_{\hat{F}(\mu_0,\gamma)}^{F_H} \left( B(Q(F), E(F), F) - \phi E(F) \right) w(F) dF. \tag{3}$$

The FOCs are given by (4) and (5) below, where the notation suppresses the dependence of $B_Q$, $B_E$, $dQ/d\mu_0$, $dQ/d\gamma$, $dE/d\mu_0$, and $dE/d\gamma$ on $\mu_0$, $\gamma$, and $F$:

$$\text{FOC}_{\mu_0} : \int_{F_L}^{\hat{F}(\mu_0, \gamma)} \left( B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) w(F) dF = 0; \tag{4}$$

$$\text{FOC}_{\gamma} : \int_{F_L}^{\hat{F}(\mu_0, \gamma)} \left( B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) w(F) dF = 0. \tag{5}$$

I show in appendix C that FOCs (4) and (5) reduce to:

$$\text{FOC}_{\mu_0} : \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{F_L}^{\hat{F}(\mu_0, \gamma)} \left( \phi + B_{EF}(\hat{F}(\mu_0, \gamma) - F) \right) w(F) dF = 0; \tag{6}$$

$$\text{FOC}_{\gamma} : \frac{\gamma W(\hat{F})(B_{QE}\gamma + B_{QQ})}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2} (\phi^2 - B_{EF}^2 \sigma_{Fc}^2) = 0, \tag{7}$$

where $\sigma_{Fc}^2$ is the variance of $F$ conditional on the standard binding.[12]

$\text{FOC}_{\mu_0}$ implies that the optimal value of $\mu_0$ should be such that the fuel price $\hat{F}$ at which the standard just binds is invariant to $\gamma$. Thus, the standards in panels (b), (c), and (d) of figure 1 all intersect agents' unrestricted choice line at the same "pivot point" $(Q^*, \mu^*)$.

$\text{FOC}_{\gamma}$ is solved by setting $\gamma = 0$ so that the optimal standard, when there is only uncertainty in $F$, is flat rather than output-based.[13] The intuition for this result flows from the fact that output-based regulation's welfare effects stemming from both the distortion to $Q$ and the flexibility in $E$ are second-order in $\gamma$. The argument for why the distortion to $Q$ is second-order is standard: output-based regulation's implicit subsidy to $Q$ creates a Harberger deadweight loss triangle (Ito and Sallee 2018). The flexibility benefit is also second-order because: (1) the response of $Q$ to shocks to $F$ is proportional to $\gamma$ (as illustrated in figure 1, panels (c) and (d)); and (2) changes in $E$ are given by the product of $\gamma$ with changes in $Q$. Thus, $dE/dF$ is proportional to $\gamma^2$ (also see equation (29) in appendix A).

---

[12]That is, $\sigma_{Fc}^2 \equiv \text{Var}(F | F \leq \hat{F}(\mu^*))$.

[13]$\gamma = -B_{QQ}/B_{QE}$ is also a solution to $\text{FOC}_{\gamma}$. Kellogg (2018) shows, however, that only the $\gamma = 0$ solution satisfies the SOC, so long as $\mu_0$ is chosen optimally (guaranteeing that $\phi^2 > B_{EF}^2 \sigma_{Fc}^2$).

## 3.2  Optimal standards under uncertainty in $\eta$

I now consider optimal regulation under uncertainty in the marginal value of $Q$, denoted by $\eta$. This subsection assumes that $F$ is fixed and known by the regulator; section 3.3 will consider uncertainty in both $\eta$ and $F$.

As with subsection 3.1, it is helpful to begin here by building intuition for agents' equilibrium choices of $Q$ and $E$ in the absence of regulation (the formal comparative statics are ensconced in appendix A). These choices are illustrated in figure 2, panel (a). Both $Q$ and $E$ increase with $\eta$, so that the line denoting agents' private choices is upward sloping in $Q$-$E$ space. However, this line has a flatter slope than was the case under uncertainty in $F$ (figure 1); this result is a consequence of the fact that the uncertainty now directly affects $Q$ rather than $E$.[14]

Under uncertainty in $F$, internalization of the externality $\phi$ was isomorphic to an increase in $F$. This relationship does not hold for $\eta$, so that the line denoting socially optimal $Q$ and $E$ now sits below rather coincident with the line denoting agents' unconstrained choices, as shown in figure 2, panel (a).

Panel (b) of figure 2 illustrates an optimal flat standard, $\mu^*$. The FOC that governs this standard, given by equation (8) below, is similar to that of equation (2), which applied when $F$ was uncertain. $\hat{\eta}$ denotes the value of $\eta$ at which the standard just binds:

$$\frac{1}{1 - V(\hat{\eta}(\mu^*))} \int_{\hat{\eta}(\mu^*)}^{\eta_H} B_E(Q(\mu^*, \eta), \mu^*, \eta) v(\eta) d\eta = \phi. \tag{8}$$

In panel (b) of figure 1, agents' choice of $Q$ was fixed at $Q^*$ whenever the standard bound, regardless of the realization of $F$. In figure 2, panel (b), however, $Q$ varies with $\eta$ when the standard is binding. Intuitively, increases in $\eta$ (the marginal value of $Q$) lead agents to increase $Q$ in equilibrium, holding $E$ fixed at $\mu^*$. This variation in $Q$ is important because

---

[14]The slope $dE/dQ$ induced by variation in $\eta$ is $-B_{QE}/B_{EE}$, while the slope induced by variation in $F$ is $-B_{QQ}/B_{QE}$. The former slope is smaller assuming that the SOC for agents' interior private optimum, $B_{QQ}B_{EE} - B_{QE}^2 > 0$, holds.

**Figure 2:** Agents' choices, social optima, and regulations under uncertainty in $\eta$
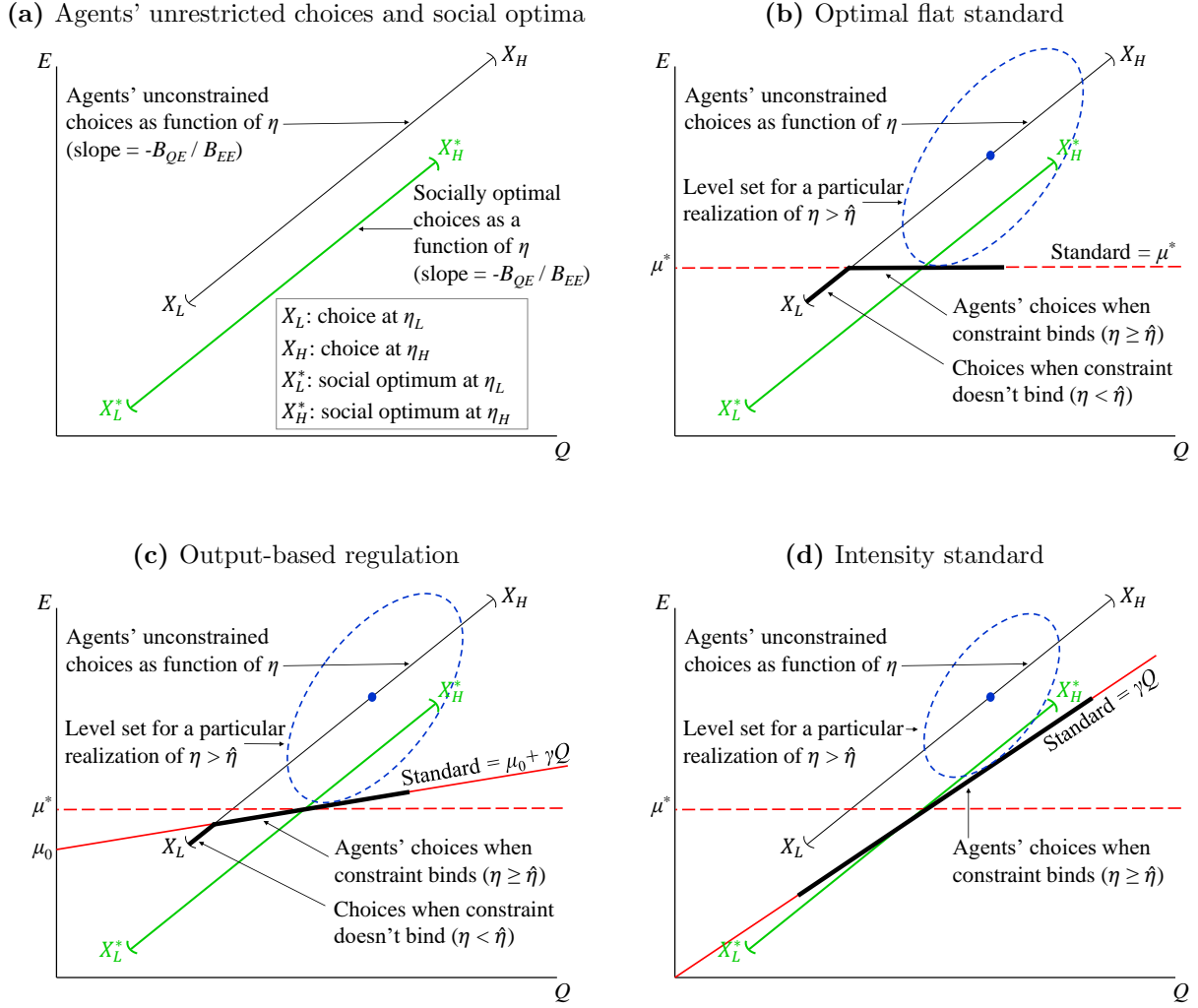
**(a)** Agents' unrestricted choices and social optima

**(b)** Optimal flat standard

**(c)** Output-based regulation

**(d)** Intensity standard

Note: $Q$ denotes the good's output (or an attribute), and $E$ denotes emissions. When unconstrained by regulation, agents' choices lie on the thin upward-sloping black line between the points $X_L$ and $X_H$, where a higher realization of $\eta$ leads to higher choices of $Q$ and $E$. The socially optimal set of outcomes is given by the lower, parallel green line, so that socially optimal choices lie between $X_L^*$ and $X_H^*$. $\eta_L$ and $\eta_H$ denote the lowest and highest possible realized net demands for $Q$, respectively. The output-based standard $\mu(Q) = \mu_0 + \gamma Q$ binds only for values of $\eta$ greater than $\hat{\eta}(\mu_0, \gamma)$. See text for details.

it implies that under output-based regulation, variation in $E$ will be first-order rather than second-order in $\gamma$. This fact gives rise to the main result of this section of the paper: under uncertainty in $\eta$, the optimal value of $\gamma$ strictly exceeds zero.

The regulator's FOCs for $\mu_0$ and $\gamma$ are given by equations (9) and (10) (suppressing the

dependence of $B_Q$, $B_E$, $dQ/d\mu_0$, $dQ/d\gamma$, $dE/d\mu_0$, and $dE/d\gamma$ on $\mu_0$, $\gamma$, and $\eta$):

$$\text{FOC}_{\mu_0} : \int_{\hat{\eta}(\mu_0,\gamma)}^{\eta_H} \left( B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) v(\eta)d\eta = 0; \tag{9}$$

$$\text{FOC}_{\gamma} : \int_{\hat{\eta}(\mu_0,\gamma)}^{\eta_H} \left( B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) v(\eta)d\eta = 0. \tag{10}$$

I show in appendix D that FOCs (9) and (10) reduce to equations (11) and (12) below, where $\sigma_{\eta c}^2$ denotes the variance of $\eta$ conditional on the standard binding:

$$\text{FOC}_{\mu_0} : \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{\hat{\eta}(\mu_0,\gamma)}^{\eta_H} \left( \frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}} B_{Q\eta}(\eta - \hat{\eta}(\mu_0,\gamma)) + \phi \right) v(\eta)d\eta = 0; \tag{11}$$

$$\text{FOC}_{\gamma} : \frac{(1 - V(\hat{\eta}(\mu_0,\gamma)))(B_{QE}\gamma + B_{QQ})\phi^2}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2} \left[ \frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}} \cdot \frac{B_{Q\eta}^2 \sigma_{\eta c}^2}{\phi^2} + \gamma \right] = 0. \tag{12}$$

FOC$_\gamma$'s solution requires a unique $\gamma \in (0, -B_{QE}/B_{EE})$.[15] The optimal standard is therefore output-based, though the optimal slope $\gamma^*$ is always strictly less than the slope of the line denoting agents' unconstrained choices in $(Q, E)$ space. The intuition that the flexibility benefits of output-based regulation are now first-order is expressed formally by the term that is not proportional to $\gamma$ inside the brackets in equation (12).

An optimal output-based standard is depicted in panel (c) of figure 2. Visually, this output-based standard does a better job of of minimizing the average (over $v(\eta)$) distance between agents' induced choices and the social optima than does the flat standard in panel (b). Also, this standard intersects agents' unconstrained choice line at a lower point than does the flat standard. This intersection is governed by FOC$_{\mu_0}$ (equation (11)), which requires that $\hat{\eta}$ decrease as $\gamma$ increases.

---

[15]For $\gamma \in [0, -B_{QE}/B_{EE}]$, the term in FOC (12) to the left of the brackets is strictly negative due to the $B_{QE}\gamma + B_{QQ}$ term and to agents' private SOC, $B_{QQ}B_{EE} - B_{QE}^2 > 0$. Inside the brackets, $(B_{QE} + B_{EE}\gamma)/(B_{QE}\gamma + B_{QQ})$ is strictly negative on $[0, -B_{QE}/B_{EE})$ and approaches zero as $\gamma \to -B_{QE}/B_{EE}$. Thus, the left hand side of FOC$_\gamma$ is strictly positive at $\gamma = 0$ and strictly negative at $\gamma = -B_{QE}/B_{EE}$, implying an interior optimum. Furthermore, the optimal $\gamma$ is unique, since the derivative of the term in brackets with respect to $\gamma$ is strictly positive for $\gamma \in (0, -B_{QE}/B_{EE})$.

Figure 2 and equation (12) together also provide guidance on when a large degree of output-basing (i.e., a large value of $\gamma$) will be welfare-maximizing. First, the optimal $\gamma$, which I denote $\gamma^*$, increases with the ratio of the variance of output (driven by the $B_{Q\eta}^2 \sigma_{\eta c}^2$ term) to the externality $\phi$. Intuitively, output-basing is valuable when the uncertainty faced by the regulator is economically large. Second, $\gamma^*$ is large when the rate at which unconstrained agents vary $E$ relative to $Q$ in response to shocks to $\eta$ (i.e., the slope $-B_{QE}/B_{EE}$) is large, since in this case output shocks can generate large changes in marginal abatement cost under a flat standard. Third, $\gamma^*$ increases with the insensitivity of agents' choice of $Q$ to changes in the cost of $Q$ (conditional on the variance of output), since an inelastic net demand for $Q$ implies small welfare losses from the distortion to $Q$ induced by output-based regulation.[16]

Finally, it is possible that $\gamma^*$ is so large that an intensity standard, which fixes $\mu_0 = 0$, yields greater expected welfare than does a flat standard. For instance, figure 2, panel (d) depicts an intensity standard that at least visually yields choices of $Q$ and $E$ that are closer to the first-best than does a flat standard.

## 3.3   Uncertainty in both $\eta$ and $F$

Finally, consider optimal output-basing when the regulator faces uncertainty in both $\eta$ and $F$. I show in appendix E that the sign of $\gamma^*$ is given by equation (13), assuming that the uniqueness of the solution to the regulator's FOCs continues to hold:

$$\text{sign}\,\gamma^* = \text{sign}\left(E_c[(\eta - E_c[\eta])^2] - \frac{B_{QQ}B_{EF}}{B_{QE}B_{Q\eta}}E_c[(F - \hat{F}(E_c[\eta]))(\eta - E_c[\eta])]\right), \qquad (13)$$

where $\hat{F}(\eta)$ denotes the fuel price at which the standard just binds, given $\eta$, and $E_c$ denotes an expectation conditional on the standard binding.

---

[16]To see this last result, note that the variance of output $Q$ under the standard is equal to $B_{Q\eta}^2 \sigma_{\eta c}^2 / (B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2$, using equation (26) from appendix A. The magnitude of the response of $Q$ to shocks to the cost of (or net demand for) $Q$ decreases with the magnitude of $B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$. Multiply and divide the left term in brackets in equation (12) by $(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2$ to see that, holding the variance of output fixed, the welfare-maximizing $\gamma$ increases with the magnitude of $B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$.

If $F$ is uncorrelated with $\eta$, then equation (13) indicates that sign $\gamma^* > 0$, as was the case in equation (12) with no uncertainty in $F$. If the covariance between $\eta$ and $F$ is positive and sufficiently large relative to the variance of $\eta$, however, equation (13) shows that the optimal value of $\gamma$ may be zero or even negative. Moving in the other direction, a negative correlation between $\eta$ and $F$ implies an increase in $\gamma^*$ relative to the case when they are uncorrelated.

The intuition behind these results flows from the fact that when $F$ is correlated with $\eta$, an output-based standard provides a means for the regulator to index the standard to $F$. That is, variation in output becomes a proxy for variation in $F$, so that an output-based standard adds value by dampening variation in marginal abatement cost induced by shocks to $F$. For instance, when $\eta$ and $F$ are negatively correlated, positive shocks to $\eta$ are associated with low fuel prices and therefore high marginal abatement costs. Expected welfare is therefore increased by allowing agents to emit more pollution when there is high demand for output, so that the optimal standard is heavily output-based. The reverse intuition holds when $\eta$ and $F$ are positively correlated. These results are related to the fact that, in general, indexing the emissions standard to any variable that is correlated with $F$ (or to $F$ itself, as discussed in Kellogg (2018)) will improve expected welfare.

# 4   Output and attribute-based standards in three settings

Section 3 presented a model showing that output or attribute-based emissions standards can increase expected welfare, relative to a flat standard, in the presence of uncertainty about agents' future marginal valuation of output (or attributes). This section evaluates the quantitative importance of this result by calibrating the model to reflect three U.S. policy settings: (1) an attribute-based fuel economy standard (holding vehicle miles traveled fixed); (2) a fuel economy standard in the presence of endogenously-chosen miles traveled (holding

**Table 1:** Interpretation of key variables in each application of the model

| Application | $Q$ | $E$ | $\eta$ | $F$ |
|---|---|---|---|---|
| Footprint-based fuel economy standards | Vehicle footprint ($ft^2$) | Gallons per 100 miles | Marginal value of footprint ($\$/ft^2$) | Gasoline price ($\$/$gal) |
| Fuel economy standard with endogenous miles traveled | Lifetime 1000s of miles traveled per vehicle | Lifetime gallons consumed per vehicle | Marginal value of miles traveled ($\$/$1000 miles) | Gasoline price ($\$/$gal) |
| Output-based $CO_2$ regulation of electricity generation | Electricity generation (TWh/yr) | $CO_2$ emissions (Mmtons/yr) | Marginal value of electricity ($\$/$TWh/yr) | Negative of natural gas price ($\$/$mmBtu) |

Units: $ft^2$ = square feet; gal = gallons; TWh/yr = Terawatt hours per year; Mmtons/yr = millions of metric tons per year; mmBtu = millions of British Thermal Units.

vehicle attributes fixed); and (3) an output-based $CO_2$ standard for electric generation.[17] Table 1 summarizes the interpretation of the variables $Q$, $E$, $\eta$, and $F$ in each application. In each case, I use the model to evaluate the expected welfare obtained from both a flat standard and an intensity standard, and I then compute the welfare-maximizing output or attribute-based standard.

In all three settings, I apply the modeling framework from sections 2 and 3 directly, abstracting away from heterogeneity or other institutional features, and maintaining the second-order Taylor approximation for $B(Q, E, \eta, F)$. Consequently, the quantitative exercises below should not be viewed as providing definitive answers but rather as providing diagnostics for when heavily output or attribute-based standards might improve expected welfare. These results can therefore guide future research that would more precisely model the relevant agents and institutions in specific settings. In addition, the fidelity of these exercises to the model helps to tightly link the results to the core intuition underlying the model and to the comparative statics discussed at the end of section 3.2.

In each setting, I allow for uncertainty in both $\eta$ and $F$, even though volatility in $F$

---

[17]Attribute-based fuel economy standards are properly modeled as an intensity standard (on fuel use as a function of miles driven) that is in turn attribute-based. Modeling the dependence of a fuel use standard on both miles driven and vehicle attributes simultaneously is beyond the scope of this paper; I therefore model the mileage-based and footprint-based aspects of U.S. fuel economy standards one at a time rather than jointly.

cannot on its own motivate output or attribute-based standards. Incorporating uncertainty in $F$ does, however, highlight the substantial difference between the first-best outcome and outcomes under emissions standards (output-based or not) when fuel price uncertainty is large, as it is in all three applications.[18] I assume that the correlation between $\eta$ and $F$ is zero, so that the welfare effects of output-basing are driven solely by the intuition presented in section 3.2 rather than by the possibility that shocks to output act as a proxy for shocks to fuel prices, as discussed in section 3.3. I do so for three reasons. First, my empirical estimates of the correlation between $\eta$ and $F$ are imprecise because only limited time series data are available to measure fluctuations in $\eta$. Yet at the same time, I find that even modest correlation between $\eta$ and $F$ greatly affects the optimal amount of output-basing, since expected welfare can be substantially increased by indexing the standard to $\eta$ when it is correlated with $F$.[19] Second, estimates using longer time series suggest that the correlation between broad economic shocks—the primary driver of $\eta$—and fuel price shocks is small. Anderson, Kellogg, and Salant (2018), for instance, finds that the CAPM beta for the price of crude oil is only 0.05, using monthly data from 1983 to 2015. Third, in practice I am not aware of any emissions standard that is indexed directly to fuel prices. Thus, it seems appropriate to eliminate this potential motivation for output-based standards here.

In each of the three applications, quantifying welfare effects requires estimates of the parameters governing $B(Q, E, \eta, F)$, the externality $\phi$, the uncertainty in $\eta$ and $F$ faced by the regulator, and $\bar{Q}$ and $\bar{E}$—the values of $Q$ and $E$ that would be chosen in the absence of regulation at the expected values of $\eta$ and $F$. To calibrate these values, I use estimates

---

[18]When I run the models with a non-stochastic and constant fuel price, the welfare rankings of flat versus intensity standards and the optimal output-based slopes $\gamma$ are qualitatively unchanged from those presented below. The optimal slopes are 0.0081 gal/100mi/ft$^2$, 0.68 gal/100mi, and 0.36 Mmtons $CO_2$/TWh in the footprint, miles traveled, and electricity applications, respectively. The expected welfare associated with these optimal slopes is substantially larger, however, than what is presented below: \$68/vehicle, \$68/vehicle, and \$2.8 billion/year.

[19]For instance, to estimate the correlation between $\eta$ and $F$ for the footprint-based standards application in section 4.1, there exist only 14 observations of year-to-year first differences in footprints and the price of gasoline. The estimated correlation coefficient is -0.34 but with a p-value of 0.23 (assuming spherical disturbances). However, this correlation is large enough in magnitude to imply that a fuel economy standard that is linear (not just affine) in footprint yields greater expected welfare than a flat standard, a result markedly different than that given in figure 3 below.

from previous studies and my own estimates of the historic volatility of attributes, output, and fuel prices. Where there are a range of plausible parameters that could be used, I select parameters that favor more output or attribute-basing, since I am interested in assessing whether substantially output or attribute-based standards might plausibly yield greater expected welfare than a flat standard. In particular, I will choose approaches that result in large estimates of the long-run uncertainty in $\eta$.

## 4.1 Footprint-based fuel economy standards

I first consider U.S. footprint-based fuel economy standards, holding vehicle miles traveled fixed. As shown in table 1, this model defines $Q$ as vehicle footprint in square feet (ft$^2$) and $E$ as vehicle fuel economy in gallons per 100 miles (gal/100mi). I measure private welfare $B(Q, E, \eta, F)$ as $ per vehicle, normalized to zero at agents' unconstrained choices at each possible realization of $\eta$ and $F$.

Table 2 presents the calibrated parameter values for this application. My estimate of $B_{QQ}$ comes from Ito and Sallee (2018), which estimates this parameter using changes in the weight of Japanese vehicles following changes in Japan's weight-based fuel economy subsidies in 2009 (I am not aware of any papers that estimate this parameter in the U.S. market). I convert that paper's estimate of $B_{QQ}$ in $ per kg$^2$ to $ per ft$^2$ using engineering estimates from Whitefoot and Skerlos (2012).[20] I obtain $B_{EE}$ from National Research Council (2015), which estimates "pathways" by which fuel economy can be improved via sequential addition of fuel-saving technology to a baseline vehicle.[21] And I derive $B_{QE}$ from the fact that the actual slope parameter $\gamma$ for U.S. fuel economy regulations was drawn to match $-B_{QE}/B_{EE}$,

---

[20]Table 3, column (2) in Ito and Sallee (2018) provides an estimate of $B_{QQ}$ of -$0.230/kg$^2$. Whitefoot and Skerlos (2012) finds that a 1 log point increase in footprint is associated with an increase of 0.53 to 1.31 log points in weight, depending on specification. I roughly split the difference and assume that weight and footprint are 1:1 in logs. Finally, data provided by Leard, Linn, and McConnell (2017) indicate that in 2012, the average U.S. vehicle had a footprint of 56.9ft$^2$ and weighed 3,666lb. Multiplying Ito and Sallee (2018)'s -$0.230/kg$^2$ by the square of this weight-to-footprint ratio yields the $B_{QQ}$ of -$197/ft$^4$ in table 2.

[21]I follow the procedure discussed in Kellogg (2018) to derive $B_{EE} = -\$1,756$ per (gallon/100 miles)$^2$ from National Research Council (2015).

**Table 2:** Calibrated parameters for U.S. footprint-based fuel economy standards

| Parameter | Value | Main sources |
|---|---|---|
| $B_{QQ}$ | -\$197 / ft$^4$ | Ito and Sallee (2018), table 3, column (2) |
| $B_{EE}$ | -\$1756 / (gal/100mi)$^2$ | National Research Council (2015), tables 8.4a and 8.4b |
| $B_{QE}$ | \$77.9 / (gal·ft$^2$/100mi) | $B_{EE}$ times slope of U.S. footprint-based standard |
| $B_{EF}$ | -115644 miles | Discounted vehicle lifetime from Busse *et al.* (2013) |
| $B_{Q\eta}$ | 1 | normalization |
| $\sigma_F$ | \$0.35 / gal | historic volatility of gasoline prices |
| $\sigma_\eta$ | \$76.83 / ft$^2$ | historic volatility of vehicle footprints |
| $\phi$ | \$486 / (gal/100mi) | Kellogg (2018) externality of \$0.42/gal, converted to vehicle lifetime using Busse *et al.* (2013) |

the rate at which private agents increase $E$ with $Q$.[22]

$B_{EF}$ in this application represents the (negative) lifetime discounted miles traveled for new vehicles. I derive $B_{EF}$ = -115,644 miles using data from Busse, Knittel, and Zettelmeyer (2013) on fleet-average scrappage probabilities and miles traveled for new vehicles, along with a discount rate of 6.2% from Allcott and Wozny (2014).[23]

I capture uncertainty in $\eta$ and $F$ from the perspective of a regulator setting an emissions standard that will apply for 10 years into the future. This approach is aligned with actual U.S. fuel economy policy, as the standards set in 2012 will apply through at least 2021 and possibly 2025 (see Kellogg (2018) for a discussion). I assume that $\eta$ and $F$ are normally distributed, with the mean of $F$ equal to the average retail gasoline price prevailing in 2012 of \$3.69/gallon.[24,25] The expected value of $\eta$ is normalized to zero.

---

[22]The original footprint-based regulation (71 FR 17565) states on p.17596 that "the agency adds fuel saving technologies to each manufacturer's fleet until the incremental cost of improving its fuel economy further just equals the incremental value of fuel savings and other benefits from doing so" for each vehicle model, and then statistically fits a line to the resulting relationship between fuel use and footprint. This slope from this line is exactly what is described by $-B_{QE}/B_{EE}$. The average slopes for 2017–2025 vehicles in the final rule (77 FR 62623, p.62782) are 3.87g $CO_2$ per ft$^2$ for cars and 4.04g $CO_2$ per ft$^2$ for trucks. Averaging these two values, applying the gasoline emissions factor of 8.91 kg $CO_2$ per gallon (EIA 2011), and multiplying by $B_{EE}$ yields the \$77.9 / (gal·ft$^2$/100mi) in table 2.

[23]I use the new vehicle data from Busse, Knittel, and Zettelmeyer (2013) that were originally sourced from the National Highway Transportation Survey.

[24]I use the tax-inclusive all grades and all formulations retail gasoline price series from the EIA (available at https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_a.htm), deflated by the Bureau of Labor Statistic's CPI for all urban consumers, all items less energy, not seasonally adjusted (series CUUR0000SA0LE), using 2012 as the base year.

[25]A no-change forecast assumption is consistent with results on the relative accuracy of no-change forecasts of the long-run real price of oil (Alquist, Kilian, and Vigfusson 2013) and relates to evidence that consumers hold no-change beliefs about future gasoline prices (Anderson, Kellogg, and Sallee 2013).

To obtain an estimate of $\sigma_F$, I estimate, over the ten-year regulatory horizon, the volatility of the three-year moving average of historic monthly U.S. gasoline prices.[26] I use the moving average rather than the raw monthly data because automakers can typically only adjust their vehicles' fuel economy during a model "refresh", and a complete refresh cycle usually takes three or four years.[27] For each month $t \in [1, 120]$, I compute the historic volatility of this moving average by taking the standard deviation of $t$-month differences in the moving average.[28] Volatility increases with the time horizon: the one-month volatility is \$0.02/gallon, while the ten-year volatility is \$0.47/gallon. I set $\sigma_F$ equal to the average volatility across all horizons $t \in [1, 120]$, yielding $\sigma_F = $ \$0.35/gallon.[29]

I identify $\sigma_\eta$ using the fact that under a flat fuel economy standard, the standard deviation of $Q$ is equal to $B_{Q\eta}\sigma_\eta / B_{QQ}$ (see equation (26) in appendix A). I therefore adopt the normalization $B_{Q\eta} = 1$ and obtain an estimate of $\sigma_\eta$ by estimating the volatility of vehicle footprints, $Q$, over a ten-year horizon, using historic data on average U.S. vehicle footprints from Leard, Linn, and McConnell (2017). Because these data are annual and only go back to 1996, it is not practical to estimate $\sigma_\eta$ by taking long differences, as I did for $\sigma_F$.[30] Instead, I estimate one-year volatility in $Q$ and then extrapolate that volatility over the ten-year policy horizon under an assumption that $\eta$ follows a random walk.[31] This assumption will yield larger estimates of volatility over long time horizons than would a model of mean-reversion. Finally, I obtain $\sigma_\eta$ by averaging volatility over the 10 years.

---

[26]The three-year moving average in each month $t$ uses data from $t$ and from the prior 35 months.

[27]Were I to instead use the raw monthly data, I would instead estimate $\sigma_F = $ \$0.65, in which case even the optimally-set attribute-based standard would achieve a welfare gain of only \$7 per vehicle. This optimal standard is essentially flat, similar to the results below.

[28]I use differences from January, 2004—the first month for which a ten-year difference can be calculated—through 2012.

[29]In principle, the calculations below of the optimal standards and their welfare outcomes should integrate over welfare effects for each month in which the standard is applied, rather than be calculated from the average volatility over this time. As in Kellogg (2018), I use the latter approach here because it is easier to implement and because it more tightly connects to the model presented in sections 2 and 3.

[30]In addition, I only use data through 2010, since after that time fuel economy standards for cars became attribute based and started increasing in stringency.

[31]Specifically, I take one-year differences in $Q$ and obtain the one-year volatility of $Q$ by estimating the standard deviation of these differences. I then extrapolate this volatility for each policy year $t$, $t \in [1, 10]$, by multiplying by $\sqrt{t}$.

To obtain $\phi$, I begin with the externality of \$0.42/gal used in Kellogg (2018), which combines the social cost of carbon from Interagency Working Group (2013) with an estimate of the foreign oil dependency externality from Parry, Wells, and Harrington (2007). I convert this value to units of \$ per (gal/100mi) by multiplying by $-B_{EF}$.

Finally, I obtain $\bar{Q}$ and $\bar{E}$ using: (1) the U.S. average footprint and fuel economy in 2012 from Leard, Linn, and McConnell (2017); and (2) data from Leard and McConnell (2017) on the shadow value of the 2012 fuel economy constraint.[32] I can then use the model to back out the values of $\bar{Q}$ and $\bar{E}$.[33]

Given these inputs, figure 3 presents footprint, fuel economy, and welfare outcomes under various forms of fuel economy regulation. The "x" denotes $(\bar{Q}, \bar{E})$, and the $\eta$ and $F$ arrows in the top-left of the figure indicate the magnitude and direction of changes in $Q$ and $E$ induced by a one standard deviation increase in $\eta$ and $F$, respectively. The representation of agents' choices of $Q$ and $E$ when unconstrained by regulation now takes the form of an ellipse rather than a line (as in figures 1 and 2), owing to the presence of uncertainty in both $\eta$ and $F$. The area encircled by the dashed black ellipse denotes agents' unconstrained choices given realizations of $\eta$ and $F$ that account for 95% of their joint probability mass. The lower dashed green ellipse indicates socially optimal choices. With the expected welfare of unregulated agents normalized to zero, the welfare gain from a Pigouvian tax on emissions is \$68 per vehicle.

The flat, solid black line (which is mostly hidden behind the red line) in figure 3 indicates the optimal flat emissions standard. This standard binds somewhat more than half the time (it is located slightly below $(\bar{Q}, \bar{E})$) and captures welfare gains of \$33 per vehicle, just 48% of the gain under the social option. This shortfall is a consequence of the large uncertainty in $F$, which results in substantial expected variation in marginal abatement costs when the

---

[32]Leard and McConnell (2017) use information on the price of tradeable fuel economy credits from Tesla Motors' 2013 SEC 10-K filing.

[33]Let $Q_0$ and $E_0$ denote the average 2012 footprint and fuel economy, and let $\phi_0$ denote the shadow value of the actual 2012 fuel economy standard. Define $D = -(B_{EE}\gamma_0 + B_{QE})/(B_{QE}\gamma_0 + B_{QQ})$. Then $\bar{E} = E_0 - \phi_0/(B_{EE} + B_{QE}D)$, and $\bar{Q} = Q_0 + (\bar{E} - E_0)D$.

**Figure 3:** Fuel economy, vehicle footprints, and welfare under footprint-based fuel economy standards
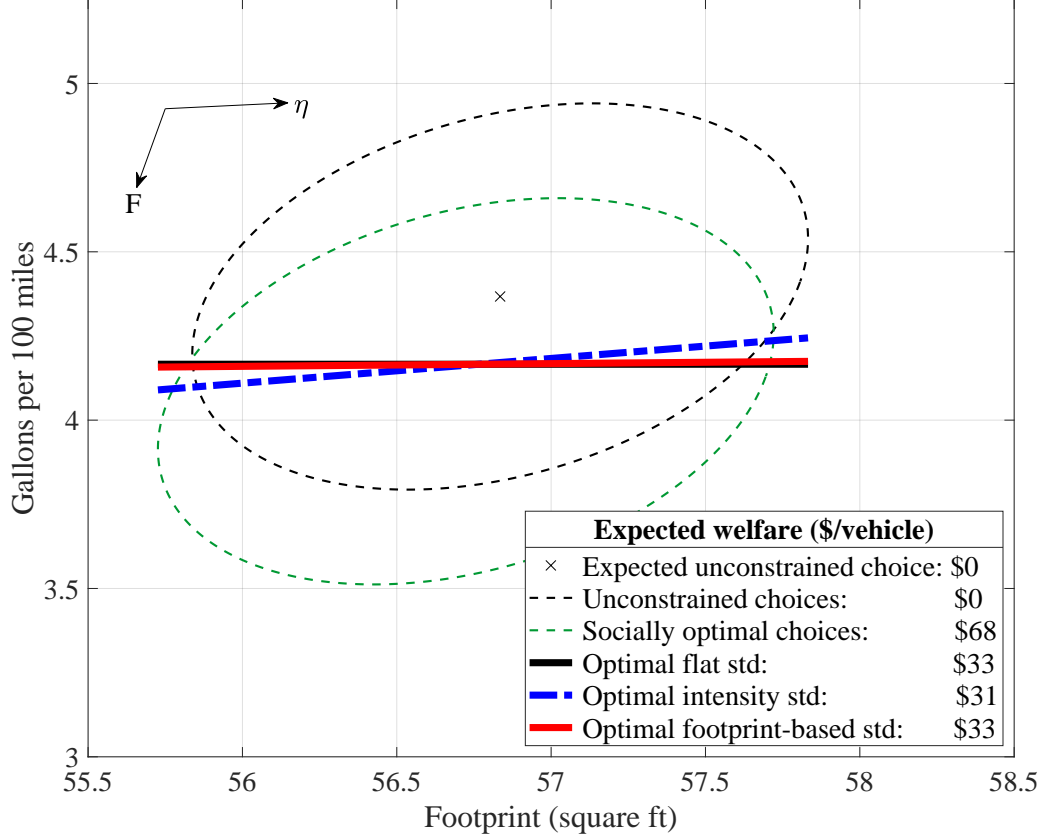


Note: Expected welfare is normalized to zero for agents' unconstrained choices. Ellipses indicate agents' choices of footprints and fuel economy associated with realizations of $\eta$ and $F$ that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The $\eta$ and $F$ arrows indicate the changes in footprint and fuel economy induced by a one standard deviation change in $\eta$ and $F$ when agents are unconstrained. See text for details.

standard binds.

The optimal amount of attribute-basing is small: the optimal value of $\gamma$ is 0.008 gallons per 100 miles per ft$^2$, equal to only 18% of the slope of the actual U.S. regulation. This optimal attribute-based standard is given by the solid red line in figure 3. In addition, the optimal intensity standard—represented by the dashed blue line—reduces welfare relative to the flat standard (and has a slope 65% greater than the actual U.S. regulation).

The limited value of attribute-based standards in this application is a consequence of

the small variance in $\eta$, the marginal value of vehicle footprint, relative to the magnitude of the externality. As can be seen in figure 3, the variance of footprint $Q$ when agents are unconstrained is quite small in percentage terms (the scale of the horizontal axis in figure 3 visually exaggerates this variance).

## 4.2 Fuel economy standards with endogenous miles traveled

In this subsection, I fix vehicle attributes and instead let $Q$ represent lifetime miles traveled per vehicle. $E$ now represents total lifetime gallons of gasoline consumed per vehicle, and I continue to measure private welfare $B(Q, E, \eta, F)$ as \$ per vehicle, normalized to zero for agents' unconstrained choices at each $F$ and $\eta$.[34]

To calibrate the values of $B_{QQ}$, $B_{QE}$, and $B_{EE}$, I first rewrite $B(Q, E, \eta, F)$ as a summation of objects that have empirical counterparts in the literature: the utility $U(Q, \eta)$ from miles traveled, fuel costs $-EF$, and the vehicle cost $-C(E/Q)$. Thus, $B(Q, E, \eta, F) = U(Q, \eta) - EF - C(E/Q)$.[35] Taking derivatives then yields:[36]

$$B_{QQ} = U_{QQ} - C'' E^2 Q^{-4} - 2C' E Q^{-3} \tag{14}$$

$$B_{QE} = C'' E Q^{-3} + C' Q^{-2} \tag{15}$$

$$B_{EE} = -C'' Q^{-2} \tag{16}$$

$U_{QQ}$ relates directly to the elasticity of miles traveled with respect to the price of gasoline $F$, holding fuel economy constant, since in that case $dQ/dF = E/(Q U_{QQ})$. I use an elasticity

---

[34]More precisely, $B(Q, E, \eta, F)$ denotes expected lifetime private welfare per vehicle at the time it is purchased new, given a no-change forecast for future gasoline prices (Anderson, Kellogg, and Sallee 2013) and demand for miles ($\eta$). Realized utility for each vehicle will be a function of realized shocks.

[35]This formulation for $B(Q, E, \eta, F)$ assumes that vehicle depreciation is entirely a function of time rather than mileage. To the extent that depreciation is a function of miles driven (and consumers account for depreciation when making driving choices), then a fuel economy policy behaves more like an emissions cap, and in fact becomes equivalent to an emissions cap in the limit in which depreciation is solely a function of mileage. I thank Mark Jacobsen for alerting me to this insight.

[36]I assume that $B(Q, E, \eta, F)$ is a second-order Taylor approximation local to the actual values of $Q_0$ and $E_0$ for new U.S. vehicles in 2012. I use $Q_0 = 115,644$ miles, consistent with the value of $-B_{EF}$ from the footprint-based regulation application in section 4.1. I calculate $E_0$ as the product of $Q_0$ with the 2012 U.S. average fuel economy of 4.01 gallons per 100 miles from Leard, Linn, and McConnell (2017).

**Table 3:** Calibrated parameters for U.S. fuel economy standards with endogenous vehicle miles traveled

| Parameter | Value | Main sources |
|---|---|---|
| $B_{QQ}$ | -\$14.96 / (1000mi)$^2$ | Gillingham (forthcoming) |
| $B_{EE}$ | -\$0.0013 / gallon$^2$ | $B_{EE}$ from table 2, scaled by vehicle lifetime miles |
| $B_{QE}$ | \$0.016 / (gal·1000mi) | $B_{EE}$ from table 2; Leard and McConnell (2017) |
| $B_{EF}$ | -1 | by construction |
| $B_{Q\eta}$ | 1 | normalization |
| $\sigma_F$ | \$0.35 / gal | historic volatility of gasoline prices |
| $\sigma_\eta$ | \$55.50 / 1000mi | historic volatility of miles traveled |
| $\phi$ | \$0.42 / gallon | Kellogg (2018) |

Note: Calibration of $B_{QQ}$, $B_{QE}$, and $B_{EE}$ also uses vehicle lifetime mileage assumptions from Busse, Knittel, and Zettelmeyer (2013) and average 2012 U.S. fuel economy from Leard, Linn, and McConnell (2017). See text for details.

of -0.081 from Gillingham (forthcoming)'s recent survey of the literature on the rebound effect.[37]

$C''$ is the same object as $B_{EE}$ from the footprint application in table 2. I calculate $C'$ using the equilibrium relationship that $-C'$ is equal to expected lifetime discounted fuel costs plus the shadow value of the fuel economy standard. I obtain the 2012 shadow value from Leard and McConnell (2017), and I calculate expected fuel costs as the product of $Q$ with the 2012 average retail gasoline price of \$3.69/gallon. Together, the estimates of $U_{QQ}$, $C'$, and $C''$ yield the values of $B_{QQ}$, $B_{QE}$, and $B_{EE}$ shown in table 3. I follow Kellogg (2018) in adopting the externality $\phi = \$0.42$ per gallon.

The calibration of $\sigma_F$ is the same as in the footprint-based standard application in section 4.1. To calibrate $\sigma_\eta$, I use the fact that volatility in miles traveled induced by volatility in $\eta$ is given by $dQ/d\eta$ times $\sigma_\eta$.[38] I obtain historical data on average U.S. vehicle miles traveled (VMT) from the Federal Reserve Bank of St. Louis.[39] Because shocks to fuel prices $F$ induce

---

[37]Of the papers on the rebound effect that Gillingham (forthcoming) surveys, -0.081 is the average elasticity among papers that use odometer readings as the primary data source.

[38]$dQ/d\eta$ is given by $-B_{Q\eta}/(B_{EE}\gamma_0^2 + 2B_{QE}\gamma_0 + B_{QQ})$, as derived for constrained agents in appendix A, equation (26), where $\gamma_0$ denotes the 2012 average fleet-wide fuel economy of 4.01 gallons per 100 miles.

[39]Annual U.S. VMT data were accessed from https://fred.stlouisfed.org/series/TRFVOLUSM227NFWA on 2 May, 2018. I use data from 1990–2010, during which time the U.S. fuel economy standard for passenger cars was constant. To convert these data from total miles traveled to miles traveled per vehicle, I scale them by multiplying by $Q_0$ (defined in footnote 36) and dividing by total miles traveled in 2012.

**Figure 4:** Gasoline consumption, miles traveled, and welfare under fuel economy standards with endogenous miles traveled
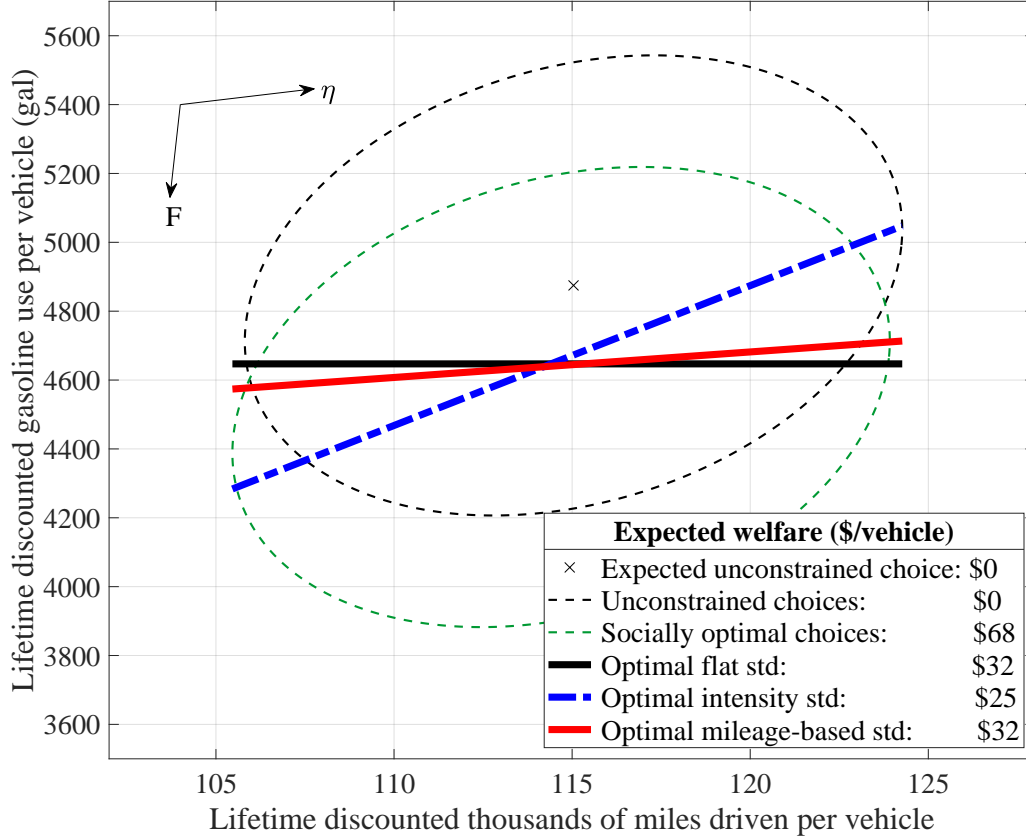


Note: Expected welfare is normalized to zero for agents' unconstrained choices. Ellipses indicate agents' choices of miles traveled and fuel economy associated with realizations of $\eta$ and $F$ that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The $\eta$ and $F$ arrows indicate the changes in miles traveled and fuel economy induced by a one standard deviation change in $\eta$ and $F$ when agents are unconstrained. See text for details.

changes to miles traveled, I must adjust the miles traveled time series for these shocks in order to isolate volatility induced by $\eta$. I do so by subtracting, in each year $t$, miles traveled equal to $dQ/dF \cdot (F_t - F_0)$, where $F_0$ denotes the 2012 gasoline price of \$3.69/gallon.[40] I then compute $\sigma_\eta$ by following the same procedure used in the footprint-based standard application in section 4.1: I compute annual volatility in miles traveled, extrapolate to a 10-year horizon

---

[40] $dQ/dF$ is given by $-B_{EF}\gamma_0/(B_{EE}\gamma_0^2 + 2B_{QE}\gamma_0 + B_{QQ})$, as derived for constrained agents in appendix A, equation (27), where $\gamma_0$ denotes the 2012 average fleet-wide fuel economy of 4.01 gallons per 100 miles.

under a random walk assumption, and then covert to $\sigma_\eta$ by dividing by $dQ/d\eta$.

I use the calibrated parameters in table 3 to generate the results in figure 4, which illustrates miles traveled, fuel economy, and welfare outcomes using the same scheme as figure 3. Relative to the footprint-based standards application, volatility in $Q$ is somewhat larger, so that the optimal standard is noticeably output-based. However, the optimal slope $\gamma$ of 0.74 gallons per 100 miles is considerably less than the slope of 4.07 gallons per 100 miles associated with the optimal intensity standard. Consequently, the optimal output-based standard achieves modestly greater expected welfare than a flat standard ($32.3 versus $32.0 per vehicle), while the welfare associated with the intensity standard is substantially less ($25.1 per vehicle). The relatively poor performance of the intensity standard is driven by: (1) the incentive it generates to increase miles traveled, even with this calibration's fairly small driving elasticity of -0.081; and (2) the fact that fuel consumption increases less than one-for-one with miles traveled in response to $\eta$ shocks, since consumers choose more efficient vehicles when they plan to drive more.

Finally, as was the case in the footprint-based results in figure 3, even the optimal output-based standard obtains less than half the expected welfare than the Pigouvian tax. This result is again a consequence of the large uncertainty in $F$.

## 4.3   Electricity sector emissions standards

Finally, I apply the model to the U.S. electricity sector. In this model, $Q$ represents electricity generation and consumption (in TWh per year) and $E$ represents $CO_2$ emissions (in Mmtons per year), as shown in table 1. The sources of uncertainty are $\eta$, representing the marginal value of electric power, and $F = -P_g$, representing the negative price of natural gas. I use the negative price because low natural gas prices increase substitution from coal to gas, which reduces rather than increases $CO_2$ emissions, holding $Q$ fixed. $B(Q, E, \eta, F)$ here denotes total annual private welfare from the U.S. electric sector. I use 2015, the year the Clean

Power Plan was finalized in the Federal Register, as the base year for the calibration.[41]

$\eta$ and $F$ experience both high and low frequency variation. Power demand, for instance, fluctuates considerably within each day but also varies across years, following economic cycles. I follow Borenstein *et al.* (forthcoming) by focusing on this lower-frequency variation when I calibrate the model, since electric sector carbon regulation typically involves long time horizons and annual (rather than daily) compliance periods.

The calibrated parameters I use are presented in table 4. To calibrate $B_{QQ}$, I use the estimated demand elasticity of -0.088 from Ito (2014), which uses residential billing data from California and incorporates responses to price lags of up to four months.[42] To calculate $B_{QQ}$, I then use the fact that $dQ/dP = 1/B_{QQ}$.[43]

To calibrate $B_{EE}$, I use Cullen and Mansur (2017)'s estimates of the effect of carbon pricing on U.S. electric sector emissions, holding electricity output fixed. That paper's estimate that a \$40 per ton tax on $CO_2$ will reduce emissions by 7.9% yields an estimate of $B_{EE}$ via the relationship $dE/d\phi = 1/B_{EE}$.[44] To obtain $B_{QE}$, I leverage the assumption that the electric sector is constant returns to scale, so that $dE/dQ = \bar{E}/\bar{Q} = 0.50$ Mmton $CO_2$ per TWh.[45] I then obtain $B_{QE}$ via the relationship $dE/dQ = -B_{QE}/B_{EE}$. Finally, I estimate $B_{EF}$ using logic from Cullen and Mansur (2017) that the effects of changes in the price of natural gas, $P_g$, can be mapped to the effects of carbon pricing, given the difference in $CO_2$ emissions between coal and natural gas.[46]

---

[41]See Federal Register (23 Oct., 2015), Vol. 80, No. 205, pp. 64661–65120.

[42]By focusing on the demand side only when calibrating $B_{QQ}$, I am implicitly assuming that electricity supply is constant returns to scale. For year-to-year variation in load, marginal generation costs may also increase with load, so that I am underestimating the total magnitude of $B_{QQ}$.

[43]I also use the value of U.S. total power consumption for 2015 of $\bar{Q} = 4,078$ TWh and the 2015 average retail price to residential end users of \$0.1265/kWh. I sourced the former value from the EIA's table of state-level generation, available at https://www.eia.gov/electricity/data/state/annual_generation_state.xls. I sourced the latter value from the EIA's 2016 Electric Power Annual Table 2.4, available at https://www.eia.gov/electricity/annual/archive/03482016.pdf.

[44]I also use $\bar{E} = 2,031$ Mmton $CO_2$ for 2015, per the EIA's 2015 Electric Power Annual, available at https://www.eia.gov/electricity/annual/html/epa_01_02.html.

[45]In comparison, Graff Zivin *et al.* (2014) estimates, on average across the U.S., that $dE/dQ = 0.55$ Mmton $CO_2$ per TWh for hourly changes in $Q$.

[46]Cullen and Mansur (2017) shows that the change in $P_g$ locally equivalent to a \$1/Mmton change in the carbon price, equivalent to $1/B_{EF}$, is given by $(CR \cdot CO_{2g} - CO_{2c})/CR$, where $CO_{2g}$ and $CO_{2c}$ are the carbon intensities of natural gas and coal, respectively, and $CR$ denotes cost ratio of coal to gas in the

**Table 4:** Calibrated parameters for U.S. electricity sector emissions

| Parameter | Value | Main sources |
| --- | --- | --- |
| $B_{QQ}$ | -\$352536 / TWh$^2$ | Ito (2014), table 3 |
| $B_{EE}$ | -\$274745 / Mmton$^2$ | Cullen and Mansur (2017), table 2 |
| $B_{QE}$ | \$136877 / (TWh·Mmton) | constant returns to scale assumption |
| $B_{EF}$ | -11.3 million mmBtu per Mmton | Cullen and Mansur (2017) and EIA |
| $B_{Q\eta}$ | 1 | normalization |
| $\sigma_F$ | \$2.59 / mmBtu | historic volatility of natural gas prices |
| $\sigma_\eta$ | \$57.0 million / TWh | historic volatility of electricity consumption |
| $\phi$ | \$38 million / Mmton | Interagency Working Group (2013) |

To calculate $\sigma_F$, I use monthly data on the front-month futures price for natural gas delivery to Henry Hub, Louisiana, which are available from the EIA back to January, 1994.[47] I focus on the price of natural gas both because it is volatile and because this price is an important driver of switches between coal and natural-gas fired generation, where natural gas involves substantially less $CO_2$ per TWh generated than coal (Cullen and Mansur 2017). I calculate $\sigma_F$ from these data using the same long-difference procedure discussed in section 4.1 to calculate $\sigma_F$ for gasoline prices, except that here I: (1) use the raw monthly data rather than a moving average; and (2) use a 15-year regulatory horizon that corresponds to the Clean Power Plan's target year of 2030.[48] I estimate $\sigma_F = \$2.59/\text{mmBtu}$.

I estimate $\sigma_\eta$ using annual data on net electricity generation from the electric power industry, dating back to 1990, available from the EIA.[49] The procedure I use is similar to that from the VMT application in section 4.2. First, I adjust the generation time series for

absence of carbon pricing. The input values for $CO_{2g}$ and $CO_{2c}$ are 117.00 and 210.86 pounds $CO_2$/mmBtu, respectively, from Cullen and Mansur (2017). To calculate $CR$, I use the 2015 average delivered coal price of \$42.58 per short ton from table 34 of the 2016 EIA Annual Coal Report, the 2015 conversion of 19.146 mmBtu per short ton from table A5 of the EIA Monthly Energy Review, and the EIA's 2015 average purchased price of natural gas of \$3.29/mmBtu (data discussed in footnote 50).

[47]These data are available at https://www.eia.gov/dnav/ng/hist/rngc1m.htm. I deflate these data to January, 2016 dollars using the CUUR0000SA0LE CPI series from the BLS. I use these data rather than purchase prices for the electric power sector (discussed in footnote 50) because those prices are available at a monthly frequency only back to 2002.

[48]I use the raw annual natural gas price data rather than a moving average because a substantial mechanism for emissions reductions from the power sector is change in utilization of coal versus gas-fired generators from the existing fleet, rather than new investment.

[49]I use national-level data that are aggregated from EIA form 923 reports. These data are available at https://www.eia.gov/electricity/data/state/annual_generation_state.xls.

changes induced by fluctuations in the price of natural gas.[50] Then, I take first (i.e., annual) differences in the adjusted generation time series, calculate annual volatility as the standard deviation of these differences, and project volatility to a 15-year horizon using a random walk assumption. I convert the average 15-year volatility of $Q$ to $\sigma_\eta$ using the fact that the volatility of $Q$ equals $dQ/d\eta$ times $\sigma_\eta$.[51] The resulting $\sigma_\eta$ equals \$57.0 million per TWh.

Finally, to calibrate $\phi$ I use the Interagency Working Group (2013)'s social cost of carbon of \$38/mton.

The model's output is presented in figure 5. For the U.S. electricity market, I find that an optimally-set intensity standard slightly outperforms an optimal flat standard, yielding expected welfare gains of \$1.42 rather than \$1.23 billion/year (relative to no regulation). The optimal output-based regulation has a slope parameter $\gamma = 0.33$ Mmtons $CO_2$ per TWh, equal to 69% of the slope of the intensity standard.

This substantial difference in results, relative to the passenger vehicle applications in sections 4.1 and 4.2, derives from two factors. First, the variance in $\eta$, the marginal value of electricity ($Q$), is relatively large in this application: $\sigma_\eta$ is equal to 4.9% of its expected value. Second, electric sector $CO_2$ emissions vary nearly linearly with output when generators are not restricted by regulation and experience demand shocks. The comparable slopes $dE/dQ$ are much flatter in the two vehicle-related applications.

The welfare ranking of the intensity standard to the flat standard is sensitive to the assumed elasticity of demand for electric power. If I use the long-run elasticity from Deryug-

---

[50]I use annual data from the EIA on the purchase price of natural gas for the electric power sector, covering 1997–2018. These data are available at https://www.eia.gov/dnav/ng/hist/n3045us3A.htm. I deflate them to January 2016 dollars using the CUUR0000SA0LE CPI series from the BLS, and I convert them from \$ per thousand cubic feet to \$/mmBtu using a conversion factor of 1.036 mmBtu/mcf from https://www.eia.gov/tools/faqs/faq.php?id=45. I also use a value of $dQ/dF$ calculated via equation (21), for unconstrained agents, in appendix A. Note that, unlike the two passenger vehicle applications, changes in the price of natural gas directly affect the marginal value of $Q$, holding $E$ fixed, since the gas price directly affects the marginal cost of power (i.e., $B_{QF} \neq 0$). Because I adjust the generation data only for the indirect effect of gas prices on $Q$, via equation (21), the direct effect is incorporated into my estimate of $\sigma_\eta$. Though this direct effect should induce a positive correlation between $F$ ($=-P_g$) and $\eta$, I estimate an overall small negative correlation of -0.07, since shocks to power demand are on average negatively correlated with $F$. In the calibration I enforce $\rho = 0$.

[51]$dQ/d\eta$ is given by $-B_{EE}B_{Q\eta}/(B_{QQ}B_{EE} - B_{QE}^2)$, as derived for unconstrained agents in appendix A, equation (20).

**Figure 5:** $CO_2$ emissions, electricity consumption, and welfare under output-based electricity emission standards
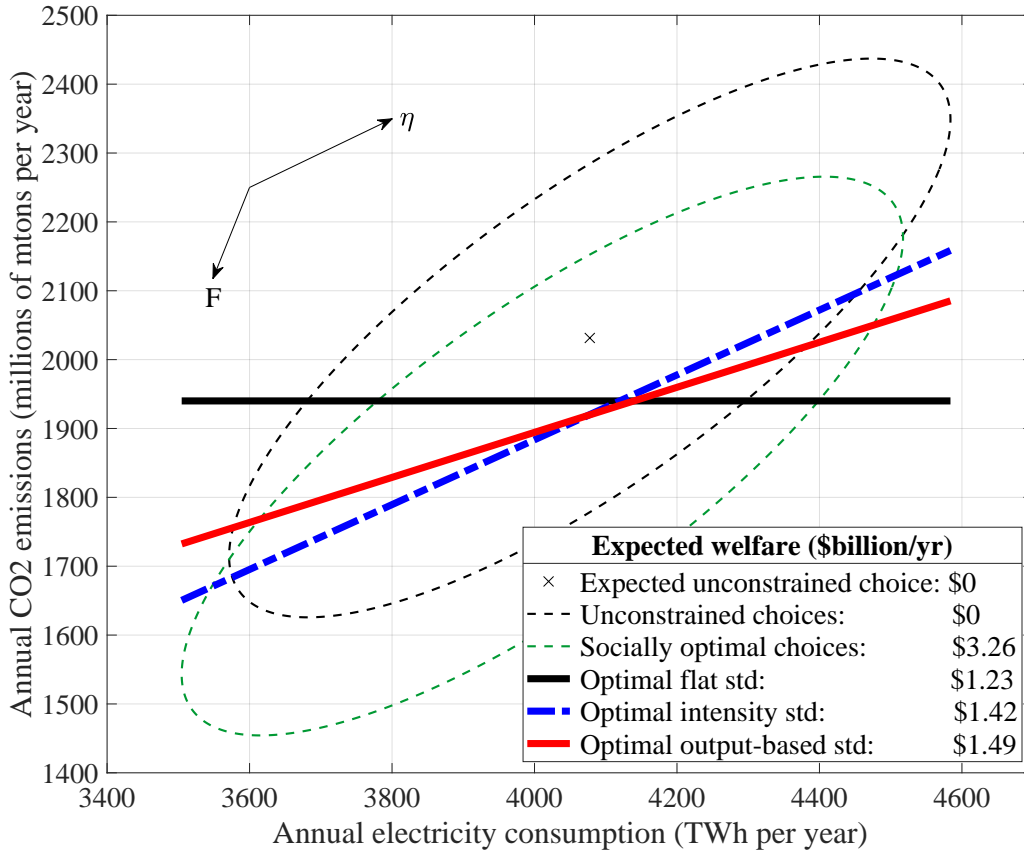


Note: Expected welfare is normalized to zero for agents' unconstrained choices. Ellipses indicate agents' choices of electricity consumption and $CO_2$ emissions associated with realizations of $\eta$ and $F$ that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The $\eta$ and $F$ arrows indicate the changes in electricity consumption and $CO_2$ emissions induced by a one standard deviation change in $\eta$ and $F$ when agents are unconstrained. See text for details.

ina, MacKay, and Reif (forthcoming) of -0.30 rather than the elasticity of -0.088 from Ito (2014), the flat standard achieves an expected welfare gain of \$2.00 billion per year, relative to \$1.69 billion per year under the intensity standard.[52] Still, even in this case I find that the welfare-maximizing standard is substantially output-based, with a slope of 0.28 Mmtons $CO_2$ per TWh.

---

[52]With a higher demand elasticity, all welfare outcomes increase. The Pigouvian tax achieves a gain of \$7.7 billion per year.

Finally, the substantial volatility of both $\eta$ and $F$ causes even the optimal output-based standard to realize substantially less expected welfare than a Pigouvian tax on emissions, which would achieve a welfare improvement of \$3.26 billion/year. The fact that uncertainty is large relative to the externality is also reflected by the fact that the flat, intensity, and optimal output-based standards all would bind only slightly more than 50% of the time. These findings are closely related to the results presented in Borenstein *et al.* (forthcoming), which studies California's cap-and-trade program and finds that uncertainty about future emissions paths is sufficiently large that the market is exceedingly likely to have an equilibrium price that lies at the administrative price floor or ceiling.

# 5 Conclusions

The consensus in the literature on pollution control is that, in the absence of clear pre-existing market distortions such as taxes, market power, or "leakage" to other sectors, $CO_2$ emissions standards that are output or attribute-based reduce welfare relative to simple, "flat" emissions limits. This paper demonstrates that this prevailing welfare result holds only when the policy-maker is certain of the demand and supply for the good(s) on which the policy is being applied. I show that in the presence of uncertainty in the marginal value of a good's output or attribute, expected welfare can always be increased by transforming a flat emissions standard into one that is output or attribute-based.

The degree to which the welfare-maximizing emissions standard depends on output or attributes is context-dependent. For U.S. attribute-based fuel economy standards, I find that the welfare-maximizing amount of attribute-basing is sufficiently small that the welfare improvement relative to a flat standard is negligible. For U.S. electricity generation, however, uncertainty about the demand for power is sufficiently large that the welfare-maximizing emissions standard is substantially output-based. Furthermore, an intensity standard may lead to greater expected welfare than a flat $CO_2$ emissions limit. This result is suggestive, as

it is sensitive to a range of plausible parameter inputs and is derived from a highly aggregated model that omits important features of the electric sector. It therefore calls for future work that would more precisely evaluate the tradeoffs between different emissions control policies in the presence of the substantial uncertainties faced by policy-makers.

Finally, it is important to not lose sight of the fact that even optimally-set output or attribute-based $CO_2$ standards fail to achieve the first-best welfare outcome in uncertain economic environments. A Pigouvian tax yields strictly greater expected welfare than any of the policies considered in this paper, as would an emissions cap that is indexed to exogenous sources of uncertainty (such as fuel prices or GDP) rather than to endogenously-determined objects such as goods' output or attributes. There is therefore a need for future research that might explain why emissions policies in practice have tended to take the form of output or attribute-based standards rather than Pigouvian taxes or exogenously-indexed emissions regulation.

# References

[1] Allcott, Hunt and Nathan Wozny (2014), "Gasoline Prices, Fuel Economy, and the Energy Paradox", *Review of Economics and Statistics* 96(5), 779-795.

[2] Alquist, Ron, Lutz Kilian, and Robert J. Vigfusson (2013), "Forecasting the Price of Oil", in Graham, Elliot and Allan Timmerman (Eds.), *Handbook of Economic Forecasting* vol.2. Amsterdam: North Holland.

[3] Anderson, Soren T., Ryan Kellogg, and Stephen W. Salant (2018), "Hotelling Under Pressure", *Journal of Political Economy* 126, 984-1026.

[4] Anderson, Soren T., Ryan Kellogg, and James M. Sallee (2013), "What Do Consumers Believe About Future Gasoline Prices?" *Journal of Environmental Economics and Management* 66, 383-403.

[5] Anderson, Soren T. and James M. Sallee (2016), "Designing Policies to Make Cars Greener: A Review of the Literature", *Annual Review of Resource Economics* 8, 157-180.

[6] Borenstein, Severin, James Bushnell, Frank A. Wolak, and Matthew Zaragoza-Watkins (forthcoming), "Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design", *American Economic Review.*

[7] 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.

[8] Cullen, Joseph A. and Erin T. Mansur (2017), "Inferring Carbon Abatement Costs in Electricity Markets: A Revealed Preference Approach Using the Shale Revolution" *American Economic Journal: Economic Policy* 9(3), 106-133.

[9] Deryugina, Tatyana, Alexander MacKay, and Julian Reif (forthcoming), "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation", *American Economic Journal: Applied Economics.*

[10] Ellerman, A. Denny and Ian Sue Wing (2003), "Absolute vs. Intensity-Based Emission Caps", MIT Joint Program on the Science and Policy of Global Change report #100.

[11] Energy Information Administration (2011), "Voluntary Reporting of Greenhouse Gases Program: Fuel Emission Coefficients". Accessed from http://www.eia.gov/oiaf/1605/coefficients.html#tbl2 on 11 April, 2016.

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

[13] Gillingham, Kenneth (forthcoming), "The Rebound Effect and the Rollback of Fuel Economy Standards", *Review of Environmental Economics and Policy.*

[14] Graff Zivin, Joshua S., Matthew J. Kotchen, and Erin T. Mansur (2014), "Spatial and Temporal Heterogeneity of Marginal Emissions: Implications for Electric Cars and Other Electricity-Shifting Policies", *Journal of Economic Behavior and Organization* 107, 248-268.

[15] Heutel, Garth (2012), "How Should Environmental Policy Respond to Business Cycles? Optimal Policy Under Persistent Productivity Shocks", *Review of Economic Dynamics* 15, 244-264.

[16] Holland, Stephen P. (2012), "Emissions Taxes Versus Intensity Standards: Second-Best Environmental Policies with Incomplete Regulation", *Journal of Environmental Economics and Management* 63, 375-387.

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

[18] Interagency Working Group on Social Cost of Carbon (2013), "Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866". United States Government. Accessed from https://www.whitehouse.gov/sites/default/files/omb/inforeg/social_cost_of_carbon_for_ria_2013_update.pdf on 19 May, 2016.

[19] Ito, Koichiro (2014), "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing" *American Economic Review* 104(2), 537-563.

[20] Ito, Koichiro and James M. Sallee (2018), "The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel Economy Standards", *Review of Economics and Statistics* 100(2), 319-336.

[21] Kellogg, Ryan (2018), "Gasoline Price Uncertainty and the Design of Fuel Economy Standards", *Journal of Public Economics* 160, 14-32.

[22] Kwoka, John E. (1983), "The Limits of Market-Oriented Regulatory Techniques: The Case of Automotive Fuel Economy", *Quarterly Journal of Economics* 98(4), 695-704.

[23] Leard, Benjamin, Joshua Linn, and Virginia McConnell (2017), "Fuel Prices, New Vehicle Fuel Economy, and Implications for Attribute-Based Standards", *Journal of the Association of Environmental and Resource Economists* 4(3), 659-700.

[24] Leard, Benjamin and Virginia McConnell (2017), "New Markets for Credit Trading under U.S. Automobile Greenhouse Gas and Fuel Economy Standards", *Review of Environmental Economics and Policy* 11(2), 207-226.

[25] Lutsey, Nic (23 Jan., 2015), "Do the automakers really need help with the U.S. efficiency standards?" International Council on Clean Transportation. Blog post. Accessed from http://www.theicct.org/blogs/staff/do-automakers-really-need-help-us-efficiency-standards on 29 Dec., 2015.

[26] National Research Council (2015), *Cost, Effectiveness, and Deployment of Fuel Economy Technologies for Light-Duty Vehicles.* Washington, D.C.: National Academy Press.

[27] Newell, Richard G. and William A. Pizer (2003), "Regulating Stock Externalities Under Uncertainty", *Journal of Environmental Economics and Management* 45, 416-432.

[28] Newell, Richard G. and William A. Pizer (2008), "Indexed Regulation", *Journal of Environmental Economics and Management* 56, 221-233.

[29] Parry, Ian W.H., Margaret Wells, and Winston Harrington (2007), "Automobile Externalities and Policies", *Journal of Economic Literature* 45(2), 373-399.

[30] Pizer, William A. and Brian Prest (2016), "Prices versus Quantities with Policy Updating", NBER working paper #22379.

[31] Quirion, Philippe (2005), "Does Uncertainty Justify Intensity Emission Caps?", *Resource and Energy Economics* 27, 343-353.

[32] 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.

[33] Weitzman, Martin L. (1974), "Prices vs. Quantities", *Review of Economic Studies* 41(4), 477-491.

[34] Whitefoot, Kate S. and Steven J. Skerlos (2012), "Design Incentives to Increase Vehicle Size Created from the U.S. Footprint-Based Fuel Economy Standards" *Energy Policy* 41, 402-411.

[35] Zhao, Jinhua (2018), "Aggregate Emission Intensity Targets: Applications to the Paris Agreement", working paper.

# Appendix

# A  Comparative statics of the model

This appendix derives the comparative statics for $dQ/d\eta$, $dQ/dF$, $dE/d\eta$, and $dE/dF$, both when agents choices' are unconstrained and when they are constrained by output-based emissions regulation.

When choices are unconstrained, agents solve the problem:

$$\max_{Q,E} B(Q, E, \eta, F). \tag{17}$$

The FOCs for this problem are given by:

$$\text{FOC}_Q : B_Q(Q, E, \eta, F) = 0 \tag{18}$$

$$\text{FOC}_E : B_E(Q, E, \eta, F) = 0. \tag{19}$$

And the SOC is $B_{QQ}B_{EE} - B_{QE}^2 > 0$. The implicit function theorem then yields the following comparative statics when agents' choices are unconstrained:

$$\frac{dQ}{d\eta} = \frac{-B_{EE}B_{Q\eta}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{20}$$

$$\frac{dQ}{dF} = \frac{B_{QE}B_{EF}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{21}$$

$$\frac{dE}{d\eta} = \frac{B_{QE}B_{Q\eta}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{22}$$

$$\frac{dE}{dF} = \frac{-B_{QQ}B_{EF}}{B_{QQ}B_{EE} - B_{QE}^2}. \tag{23}$$

When agents are constrained to the regulation $E = \mu_0 + \gamma Q$, they instead solve:

$$\max_{Q} B(Q, \mu_0 + \gamma Q, \eta, F). \tag{24}$$

The FOC for this problem, using the notation $E(Q) = \mu_0 + \gamma Q$, is given by:

$$\text{FOC}_Q : \ B_Q(Q, E(Q), \eta, F) + \gamma B_E(Q, E(Q), \eta, F) = 0. \tag{25}$$

And the SOC is $B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ} < 0$. The implicit function theorem then yields the following comparative statics when agents' choices are constrained by output-based emissions regulation (where the notation suppresses the dependence of $B_Q$, $B_E$, $dQ/d\eta$, $dQ/dF$, $dE/d\eta$,

$dE/dF$, $dQ/d\mu_0$, $dE/d\mu_0$, $dQ/d\gamma$, and $dE/d\gamma$ on $\mu_0$, $\gamma$, $\eta$, and $F$):

$$\frac{dQ}{d\eta} = \frac{-B_{Q\eta}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{26}$$

$$\frac{dQ}{dF} = \frac{-B_{EF}\gamma}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{27}$$

$$\frac{dE}{d\eta} = \frac{-B_{Q\eta}\gamma}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{28}$$

$$\frac{dE}{dF} = \frac{-B_{EF}\gamma^2}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{29}$$

$$\frac{dQ}{d\mu_0} = \frac{-(B_{EE}\gamma + B_{QE})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{30}$$

$$\frac{dE}{d\mu_0} = 1 + \gamma\frac{dQ}{d\mu_0} = \frac{B_{QE}\gamma + B_{QQ}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{31}$$

$$\frac{dQ}{d\gamma} = \frac{-(B_E + B_{EE}Q\gamma + B_{QE}Q)}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{32}$$

$$\frac{dE}{d\gamma} = Q + \gamma\frac{dQ}{d\gamma} = \frac{-B_E\gamma + B_{QE}Q\gamma + B_{QQ}Q}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}}. \tag{33}$$

# B    Representative consumer model

Consider a unit mass of agents $i$, each of whom purchases a good with quantity (or attribute) $Q_i$ and emissions $E_i$, yielding private benefit $B^i(Q_i, E_i, \eta, F)$ that can be well-approximated by a second-order Taylor expansion. This section, which closely follows Kellogg (2018), proves that a sufficient statistic for the effects of emissions standards on utilitarian social welfare is given by the private benefit function $B(Q, E, \eta, F)$ less damages $\phi E$, where $Q$ and $E$ are the sum (or average) of the $Q_i$ and $E_i$, and the first derivatives of $B$ equal the sum (or average) of the $B_Q^i$ and $B_E^i$, assuming: (1) all second derivatives of $B$ are identical across agents; (2) inclusion of compliance trading in any emissions standard; and (3) equal and constant marginal utility of income (and welfare weights) across agents.

To begin, note that with compliance trading, and given values for $\eta$ and $F$, any emissions standard is equivalent to a policy that taxes (or subsidizes) $Q$ and $E$, since all agents must face the same permit price in competitive equilibrium. Given a regulatory slope $\gamma$, $\text{FOC}_Q$ in equation (25) must hold for all agents, so that $B_Q^i = -\gamma B_E^i$ $\forall i$. And if the price of an emissions permit is given by $\tau$, we have $B_Q^i = -\gamma\tau$ and $B_E^i = \tau$ $\forall i$. Thus, given values for $\eta$ and $F$, I can model agents' welfare under an emissions standard by instead modeling a tax $\tau$ on $E$ and a tax $\tau_Q \equiv -\gamma\tau$ on $Q$.

Next, I show that imposition of a tax $\tau$ on $E$ and a tax $\tau_Q \equiv -\gamma\tau$ on $Q$ yields identical changes in $Q$ and $E$ for all agents. A given agent $i$'s FOCs under these taxes are given by:

$$\text{FOC}_Q : B_Q(Q, E, \eta, F) = \tau_Q \tag{34}$$

$$\text{FOC}_E : B_E(Q, E, \eta, F) = \tau. \tag{35}$$

Following appendix A, application of the implicit function theorem yields the following comparative statics:

$$\frac{dQ_i}{d\tau_Q} = \frac{B_{EE}}{B_{QQ}B_{EE} - B_{QE}^2}; \ \frac{dQ_i}{d\tau} = \frac{-B_{QE}}{B_{QQ}B_{EE} - B_{QE}^2}; \tag{36}$$

$$\frac{dE_i}{d\tau_Q} = \frac{-B_{QE}}{B_{QQ}B_{EE} - B_{QE}^2}; \ \frac{dE_i}{d\tau} = \frac{B_{QQ}}{B_{QQ}B_{EE} - B_{QE}^2}. \tag{37}$$

If all second derivatives of $B$ are identical for all agents, then each of the above derivatives will also be identical for all agents. Thus, emissions standards in this setting will have identical effects on each agent's choices of $Q$ and $E$.

Finally, I show that the aggregate social welfare function $\int_i (B^i(Q_i, E_i, \eta, F) - \phi E_i) di$ can be written as the sum of the representative agent's welfare function $B(Q, E, \eta, F)$ in the text (where $Q = \int_i Q_i di$ and $E = \int_i E_i di$), $-\phi E$, and terms that are policy-invariant. Let $Q_0$ and $E_0$ denote total quantity and emissions with no emissions policy and with $\eta = \bar{\eta}$ and $F = \bar{F}$ (where $\bar{\eta}$ and $\bar{F}$ denote the expected values of $\eta$ and $F$). Using a second-order Taylor approximation around $Q_0, E_0, \bar{\eta}$, and $\bar{F}$, social welfare is then given by:[53]

$$\int_i (B^i(Q_i, E_i, \eta, F) - \phi E_i) di = \int_i B_{Q0}^i (Q_i - Q_0) di + \int_i B_{E0}^i (E_i - E_0) di$$

$$+ \frac{1}{2} B_{QQ} \int_i (Q_i - Q_0)^2 di + \frac{1}{2} B_{EE} \int_i (E_i - E_0)^2 di$$

$$+ B_{QE} \int_i (Q_i - Q_0)(E_i - E_0) di + B_{Q\eta} \int_i (Q_i - Q_0)(\eta - \bar{\eta}) di$$

$$+ B_{EF} \int_i (E_i - E_0)(F - \bar{F}) di - \phi E, \tag{38}$$

where $B_{Q0}^i$ and $B_{E0}^i$ denote the values of agents' first derivatives of $B^i$ at $(Q_0, E_0, \bar{\eta}, \bar{F})$.

Replace the $(Q_i - Q_0)$ terms in equation (38) with $(Q_i - Q + Q - Q_0)$ and do likewise with the $(E_i - E_0)$ terms. Define $B_{Q0} \equiv \int_i B_{Q0}^i di$ and $B_{E0} \equiv \int_i B_{E0}^i di$. After removing terms

---

[53]The assumption of equal and constant marginal utility of income (and welfare weights) across agents is necessary for equation (38) to represent utilitarian social welfare, since in this case all wealth effects across agents cancel out in aggregate.

for which the expectation is zero, and after grouping related terms together, we obtain:

$$
\begin{aligned}
\int_i (B^i(Q_i, E_i, \eta, F) - \phi E_i) di = & [B_{Q0}(Q - Q_0) + B_{E0}(E - E_0) + \frac{1}{2} B_{QQ}(Q - Q_0)^2 \\
& + \frac{1}{2} B_{EE}(E - E_0)^2 + B_{QE}(Q - Q_0)(E - E_0) \\
& + B_{Q\eta}(Q - Q_0)(\eta - \bar{\eta}) + B_{EF}(E - E_0)(F - \bar{F}) - \phi E] \\
& + [\int_i B^i_{Q0}(Q_i - Q) di + \int_i B^i_{E0}(E_i - E) di \\
& + \frac{1}{2} B_{QQ} \int_i (Q_i - Q)^2 di + \frac{1}{2} B_{EE} \int_i (E_i - E)^2 di \\
& + B_{QE} \int_i (Q_i - Q)(E_i - E) di]
\end{aligned}
\tag{39}
$$

The first bracketed term in equation (39) is the second order Taylor expansion of $B(Q, E, \eta, F) - \phi E$. The second bracketed term is policy-invariant, since $Q_i - Q$ and $E_i - E$ are constant for each $i$. Thus, $B(Q, E, \eta, F) - \phi E$ is a sufficient statistic for the social welfare impacts of the emissions standards considered in the paper.

## C  Welfare-maximizing standard under uncertainty in $F$

This appendix derives equations (6) and (7) in the main text. The argument closely follows appendix B.3 in Kellogg (2018).

Begin by developing simple expressions for $B_E$ and $B_Q$ while the agents are constrained. Given $\mu_0$ and $\gamma$, let $\hat{F}$ denote the fuel price at which the standard just binds, and let $\hat{Q}$ and $\hat{E}$ denote the agents' choices of $Q$ and $E$ at $\hat{F}$. Because $B_Q(\hat{Q}, \hat{E}, F) = 0 \ \forall F$ (recall that $B_{QF} = 0$), we may write the Taylor expansion for $B_Q$ as:

$$
B_Q(Q, E, F) = B_{QE}(E - \hat{E}) + B_{QQ}(Q - \hat{Q})
\tag{40}
$$

On the $E = \mu_0 + \gamma Q$ standard, use equations (27) and (29) to replace $E - \hat{E}$ and $Q - \hat{Q}$ in (40), yielding:

$$
B_Q(Q, E, F) = \frac{-\gamma(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} B_{EF}(F - \hat{F}).
\tag{41}
$$

In addition, equation (25) implies that $B_E = -B_Q/\gamma$, implying:

$$
B_E(Q, E, F) = \frac{B_{QE}\gamma + B_{QQ}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} B_{EF}(F - \hat{F}).
\tag{42}
$$

Now, we can use equations (25), (30), (31), and (42) to derive equation (6) from equation (4) in the main text. Streamlining notation by defining $S \equiv B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$, we

have:

$$\text{FOC}_{\mu_0} : \int_{F_L}^{\hat{F}} \left( B_E(\frac{dE}{d\mu_0} - \gamma \frac{dQ}{d\mu_0}) - \phi \frac{dE}{d\mu_0} \right) w(F)dF = 0 \tag{43}$$

$$\Leftrightarrow \frac{1}{S} \int_{F_L}^{\hat{F}} \left( \frac{(B_{QE}\gamma + B_{QQ})B_{EF}(F - \hat{F})S}{S} - \phi(B_{QE}\gamma + B_{QQ}) \right) w(F)dF = 0 \tag{44}$$

$$\Leftrightarrow \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{F_L}^{\hat{F}} \left( \phi + B_{EF}(\hat{F} - F) \right) w(F)dF = 0. \tag{45}$$

Equation (45) matches equation (6) in the main text.

Now work with the FOC for $\gamma$, given by equation (5) in the main text. Applying equations (25), (32), (33), and (42), we obtain:

$$\text{FOC}_\gamma : \int_{F_L}^{\hat{F}} \left( B_E(\frac{dE}{d\gamma} - \gamma \frac{dQ}{d\gamma}) - \phi \frac{dE}{d\gamma} \right) w(F)dF = 0 \tag{46}$$

$$\Leftrightarrow \frac{1}{S} \int_{F_L}^{\hat{F}} \left( \frac{(B_{QE}\gamma + B_{QQ})B_{EF}(F - \hat{F})SQ}{S} - \phi(-B_E\gamma + B_{QE}Q\gamma + B_{QQ}Q) \right) w(F)dF = 0$$

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left( (B_{EF}(F - \hat{F})Q - \phi Q)S + \phi B_{EF}(F - \hat{F})\gamma \right) w(F)dF = 0. \tag{47}$$

Now use $\text{FOC}_{\mu_0}$ (equation (45)) to simplify further. In particular, use the fact that $\hat{Q} \int_{F_L}^{\hat{F}} \left( \phi + B_{EF}(\hat{F} - F) \right) w(F)dF = 0$ to transform equation (47) to:

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left( (Q - \hat{Q})(B_{EF}(F - \hat{F}) - \phi)S + \phi^2\gamma \right) w(F)dF = 0. \tag{48}$$

Then apply the derivative $dQ/dF$ (equation (27)) to eliminate the $Q - \hat{Q}$ term:

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left( -B_{EF}\gamma(F - \hat{F})(B_{EF}(F - \hat{F}) - \phi) + \phi^2\gamma \right) w(F)dF = 0. \tag{49}$$

Applying $\text{FOC}_{\mu_0}$ (equation (45)) again yields:

$$\Leftrightarrow \frac{\gamma(B_{QE}\gamma + B_{QQ})}{S^2} \int_{F_L}^{\hat{F}} \left( -B_{EF}^2(F - \hat{F})^2 + 2\phi^2 \right) w(F)dF = 0. \tag{50}$$

Define $\dot{F} \equiv \frac{1}{W(\hat{F})} \int_{F_L}^{\hat{F}} Fw(F)dF$ (i.e., $\dot{F}$ is the expected value of $F$ conditional on the standard binding). We can then simplify the $-B_{EF}^2(F - \hat{F})^2$ term inside the integral in

equation (50) as follows:

$$B_{eF}^2 \int_{F_L}^{\hat{F}} (F - \hat{F})^2 w(F) dF = B_{eF}^2 \int_{F_L}^{\hat{F}} (F - \dot{F} + \dot{F} - \hat{F})^2 w(F) dF$$

$$= W(\hat{F}) B_{eF}^2 (\sigma_{Fc}^2 + (\dot{F} - \hat{F})^2)$$

$$= W(\hat{F})(B_{eF}^2 \sigma_{Fc}^2 + \phi^2), \tag{51}$$

where the last line makes use of equation (45) and the definition of $\dot{F}$.
Finally, substitute equation (51) into equation (50) to obtain:

$$\text{FOC}_\gamma : \quad \frac{\gamma W(\hat{F})(B_{QE}\gamma + B_{QQ})}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2}(\phi^2 - B_{EF}^2 \sigma_{Fc}^2) = 0, \tag{52}$$

which matches equation (7) in the main text.

# D  Welfare-maximizing standard under uncertainty in $\eta$

This appendix derives equations (11) and (12) in the main text.

Begin by developing simple expressions for $B_E$ and $B_Q$ while the agents are constrained. Given $\mu_0$ and $\gamma$, let $\hat{\eta}$ denote the value of $\eta$ (the shock to the marginal value of $Q$) at which the standard just binds, and let $\hat{Q}$ and $\hat{E}$ denote the agents' choices of $Q$ and $E$ at $\hat{\eta}$. Because $B_E(\hat{Q}, \hat{E}, \eta) = 0 \ \forall \eta$ (recall that $B_{E\eta} = 0$), we may write the Taylor expansion for $B_E$ as:

$$B_E(Q, E, \eta) = B_{QE}(Q - \hat{Q}) + B_{EE}(E - \hat{E}) \tag{53}$$

On the $E = \mu_0 + \gamma Q$ standard, use equations (26) and (28) to replace $Q - \hat{Q}$ and $E - \hat{E}$ in (53), yielding:

$$B_E(Q, E, \eta) = \frac{-(B_{QE} + B_{EE}\gamma)}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} B_{Q\eta}(\eta - \hat{\eta}). \tag{54}$$

Now, we can use equations (25), (30), (31), and (54) to derive equation (11) from equation (9) in the main text. Streamlining notation by defining $S \equiv B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$, we

have:

$$\text{FOC}_{\mu_0} : \int_{\hat{\eta}}^{\eta_H} \left( B_E(\frac{dE}{d\mu_0} - \gamma\frac{dQ}{d\mu_0}) - \phi\frac{dE}{d\mu_0} \right) v(\eta)d\eta = 0 \tag{55}$$

$$\Leftrightarrow \frac{1}{S} \int_{\hat{\eta}}^{\eta_H} \left( \frac{-(B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta})S}{S} - \phi(B_{QE}\gamma + B_{QQ}) \right) v(\eta)d\eta = 0 \tag{56}$$

$$\Leftrightarrow \frac{-1}{S} \int_{\hat{\eta}}^{\eta_H} ((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))\, v(\eta)d\eta = 0 \tag{57}$$

$$\Leftrightarrow \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{\hat{\eta}}^{\eta_H} \left( \frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}}B_{Q\eta}(\eta - \hat{\eta}) + \phi \right) v(\eta)d\eta = 0. \tag{58}$$

Equation (58) matches equation (11) in the main text.

Now work with the FOC for $\gamma$, given by equation (10) in the main text. Applying equations (25), (32), (33), and (54), we obtain:

$$\text{FOC}_{\gamma} : \int_{\hat{\eta}}^{\eta_H} \left( B_E(\frac{dE}{d\gamma} - \gamma\frac{dQ}{d\gamma}) - \phi\frac{dE}{d\gamma} \right) v(\eta)d\eta = 0 \tag{59}$$

$$\Leftrightarrow \frac{1}{S} \int_{\hat{\eta}}^{\eta_H} \left( \frac{-(B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta})SQ}{S} - \phi(-B_E\gamma + B_{QE}Q\gamma + B_{QQ}Q) \right) v(\eta)d\eta = 0$$

$$\Leftrightarrow \frac{-1}{S^2} \int_{\hat{\eta}}^{\eta_H} (((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))QS$$
$$+ \phi\gamma(B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}))v(\eta)d\eta = 0 \tag{60}$$

Now use $\text{FOC}_{\mu_0}$ (equation (57)) to simplify further. In particular, use the fact that $\hat{Q} \int_{\hat{\eta}}^{\eta_H} ((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))\, v(\eta)d\eta = 0$ to transform equation (60) to:

$$\Leftrightarrow \frac{-1}{S^2} \int_{\hat{\eta}}^{\eta_H} (((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))(Q - \hat{Q})S$$
$$- \phi^2\gamma(B_{QE}\gamma + B_{QQ}))v(\eta)d\eta = 0 \tag{61}$$

Then apply the derivative $dQ/d\eta$ (equation (26)) to eliminate the $Q - \hat{Q}$ term:

$$\Leftrightarrow \frac{1}{S^2} \int_{\hat{\eta}}^{\eta_H} (((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))B_{Q\eta}(\eta - \hat{\eta})$$
$$+ \phi^2\gamma(B_{QE}\gamma + B_{QQ}))v(\eta)d\eta = 0 \tag{62}$$

Applying FOC$_{\mu_0}$ (equation (57)) again yields:

$$\Leftrightarrow \frac{1}{S^2} \int_{\hat{\eta}}^{\eta_H} ((B_{QE} + B_{EE}\gamma)B_{Q\eta}^2(\eta - \hat{\eta})^2 - \frac{\phi^2(B_{QE}\gamma + B_{QQ})^2}{B_{QE} + B_{EE}\gamma}$$
$$+ \phi^2\gamma(B_{QE}\gamma + B_{QQ}))v(\eta)d\eta = 0 \tag{63}$$

Define $\dot{\eta} \equiv \frac{1}{(1-V(\hat{\eta}))} \int_{\hat{\eta}}^{\eta_H} \eta v(\eta)d\eta$ (i.e., $\dot{\eta}$ is the expected value of $\eta$ conditional on the standard binding). We can then simplify the $B_{Q\eta}^2(\eta - \hat{\eta})^2$ term inside the integral in equation (63) as follows:

$$B_{Q\eta}^2 \int_{\hat{\eta}}^{\eta_H} (\eta - \hat{\eta})^2 v(\eta)d\eta = B_{Q\eta}^2 \int_{\hat{\eta}}^{\eta_H} (\eta - \dot{\eta} + \dot{\eta} - \hat{\eta})^2 v(\eta)d\eta$$
$$= (1 - V(\hat{\eta}))B_{Q\eta}^2(\sigma_{\eta c}^2 + (\dot{\eta} - \hat{\eta})^2)$$
$$= (1 - V(\hat{\eta})) \left( B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\frac{(B_{QE}\gamma + B_{QQ})^2}{(B_{QE} + B_{EE}\gamma)^2} \right) \tag{64}$$

where the last line makes use of equation (57) and the definition of $\dot{\eta}$.

Finally, substitute equation (64) into equation (63) and combine terms to obtain:

$$\text{FOC}_\gamma : \frac{1 - V(\hat{\eta})}{S^2}[(B_{QE} + B_{EE}\gamma) \left( B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\frac{(B_{QE}\gamma + B_{QQ})^2}{(B_{QE} + B_{EE}\gamma)^2} \right) - \frac{\phi^2(B_{QE}\gamma + B_{QQ})^2}{B_{QE} + B_{EE}\gamma}$$
$$+ \phi^2\gamma(B_{QE}\gamma + B_{QQ})] = 0 \tag{65}$$
$$\Leftrightarrow \frac{1 - V(\hat{\eta})}{S^2}[(B_{QE} + B_{EE}\gamma)B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\gamma(B_{QE}\gamma + B_{QQ})] = 0 \tag{66}$$
$$\Leftrightarrow \frac{(1 - V(\hat{\eta}))(B_{QE}\gamma + B_{QQ})\phi^2}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2} \left[ \frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}} \cdot \frac{B_{Q\eta}^2\sigma_{\eta c}^2}{\phi^2} + \gamma \right] = 0, \tag{67}$$

which matches equation (12) in the main text.

# E  Welfare-maximizing standard under uncertainty in $\eta$ and $F$

This appendix derives equation (13) in the main text. Under uncertainty in both $\eta$ and $F$, we can write the regulator's FOCs for $\mu_0$ and $\gamma$ as:

$$\text{FOC}_{\mu_0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \left( B_Q\frac{dQ}{d\mu_0} + B_E\frac{dE}{d\mu_0} - \phi\frac{dE}{d\mu_0} \right) w(F|\eta)v(\eta)dFd\eta = 0 \tag{68}$$

$$\text{FOC}_\gamma : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \left( B_Q\frac{dQ}{d\gamma} + B_E\frac{dE}{d\gamma} - \phi\frac{dE}{d\gamma} \right) w(F|\eta)v(\eta)dFd\eta = 0, \tag{69}$$

where $\hat{F}$ is shorthand for $\hat{F}(\mu_0, \gamma, \eta)$, the fuel price $F$ at which the standard just binds, given a regulation $(\mu_0, \gamma)$ and a demand shock $\eta$.

Because deriving an analytic expression for $\gamma^*$ is algebraically intractable, I instead evaluate the FOCs (68) and (69) at $\gamma = 0$, where the sign of $\text{FOC}_\gamma$ informs the sign of $\gamma^*$ (and is equivalent to the sign of $\gamma^*$ assuming that the uniqueness property discussed in footnote 15 holds when $F$ is stochastic).

Using expressions (30), (31), (32), and (33), FOCs (68) and (69) reduce to the following at $\gamma = 0$:

$$\text{FOC}_{\mu_0}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E - \phi)w(F|\eta)v(\eta)dFd\eta = 0 \tag{70}$$

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E Q - \phi Q)w(F|\eta)v(\eta)dFd\eta = 0. \tag{71}$$

Let $\bar{\eta} \equiv E_c[\eta]$ and let $(\bar{Q}, \bar{E})$ denote agents' choices given realizations $\bar{\eta}$ and $\hat{F}(\mu_0, \gamma, \bar{\eta})$. Use the fact from equation (70) that $\int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E - \phi)\bar{Q}w(F|\eta)v(\eta)dFd\eta = 0$ to rewrite equation (71) as:

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E - \phi)(Q - \bar{Q})w(F|\eta)v(\eta)dFd\eta = 0. \tag{72}$$

Because: (1) $E - \bar{E} = 0$ for $\gamma = 0$ (since emissions are fixed at $\bar{E}$ when the regulation binds); (2) $B_{E\eta} = 0$; and (3) $Q - \bar{Q} = -B_{Q\eta}(\eta - \bar{\eta})/B_{QQ}$ when $\gamma = 0$, we have:

$$B_E(Q, E, \eta, F) = \frac{-B_{QE}B_{Q\eta}(\eta - \bar{\eta})}{B_{QQ}} + B_{EF}(F - \hat{F}). \tag{73}$$

Substituting into equation (72) yields:

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \left( \frac{-B_{QE}B_{Q\eta}(\eta - \bar{\eta})}{B_{QQ}} + B_{EF}(F - \hat{F}) - \phi \right) \left( \frac{-B_{Q\eta}}{B_{QQ}}(\eta - \bar{\eta}) \right) w(F|\eta)v(\eta)dFd\eta = 0. \tag{74}$$

The term involving $\phi$ is zero because $E_c[\eta - \bar{\eta}] = 0$. Thus, we have:

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \frac{1}{B_{QQ}^2} \left( B_{QE}B_{Q\eta}^2(\eta - \bar{\eta})^2 - B_{QQ}B_{EF}B_{Q\eta}(\eta - \bar{\eta})(F - \hat{F}) \right) w(F|\eta)v(\eta)dFd\eta = 0, \tag{75}$$

which implies:

$$\text{sign } \text{FOC}_{\gamma}|_{\gamma=0} = \text{sign} \left( E_c[(\eta - E_c[\eta])^2] - \frac{B_{QQ}B_{EF}}{B_{QE}B_{Q\eta}} E_c[(F - \hat{F}(E_c[\eta]))(\eta - E_c[\eta])] \right) \tag{76}$$

The right-hand side of equation (76) matches the right-hand side of equation (13) in the main text. Also note that if $F$ is non-stochastic, equation (75) reduces to equation (67) (same as equation (12)) for $\gamma = 0$.