Output and Attribute-Based Carbon Regulation Under Uncertainty

Ryan Kellogg[*]

August 13, 2020

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 strictly 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.

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 take the form of output-based standards that 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. These standards often take the specific form of intensity standards that specify proportionality between emissions and output. 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.

Intensity standards generate well-understood incentives to distort output, as first recognized in Kwoka's (1983) analysis of how fuel economy standards lead to the well-known "rebound effect" that causes an increase in vehicle miles traveled. Holland, Hughes, and Knittel (2009) makes this point 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. Holland (2012) shows that the same logic extends generally to output-based standards.[1] Similar intuition also applies to settings where the emissions standard is a function of goods' attributes rather than output. For instance, Ito and Sallee (2018) shows that basing fuel economy standards on vehicles' weight or footprint results in inefficiently over-sized vehicles.

---

[1] Output-based emissions permit allocation schemes, such as that used for some industrial sectors in California, are an example of an output-based standard that is not an intensity standard.

In this paper, I ask whether output or attribute-based emissions regulation can actually deliver strict welfare improvements relative to a "flat" (non-output or attribute-based) emissions standard in the presence of uncertainty about future abatement costs. The intuition is closely related to Weitzman's (1974) prices versus quantities results. 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)—Weitzman's (1974) logic implies that a carbon tax "price" policy can achieve the first-best outcome. Under flat cap-and-trade, however, the quantity of emissions is fixed so that shocks to the economic environment cause fluctuations in marginal abatement cost. Per Weitzman (1974), this variation reduces welfare relative to the first-best. The potential virtue of output or attribute-based regulations is then that they can dampen these marginal abatement cost fluctuations and their associated welfare losses, since emissions are allowed to vary as shocks occur. Indeed, compliance flexibility was one of the justifications offered for making U.S. fuel economy standards footprint-based (see, for instance, Lutsey (2015)).

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.

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

---

[2]Pizer and Prest (forthcoming) 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.

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.

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 yields greater expected welfare than a flat standard. Still, even an optimally-set output or attribute-based standard is a second-best policy solution: a Pigouvian tax remains first-best because it maintains compliance flexibility while avoiding distortions to output or attributes.

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.* (2019), 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

---

[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.

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 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, incomplete information, or other taxes or regulations, any of which could on its own justify attribute or output-basing as a second-best policy.[4]

---

[4]Ito and Sallee (2018), for instance, includes a discussion of how other market distortions might provide an efficiency justification for attribute-based fuel economy standards, noting that vehicle accident externalities actually imply efficiency gains from taxing vehicle size (which "reverse" attribute-based fuel economy standards—in which larger vehicles face a tougher standard—would implicitly do), since accident externalities increase with vehicle size. One pre-existing distortion that I consider explicitly in appendix A.6 is the case of whether to attribute-base a standard that is already output-based, in the presence of uncertainty about

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: η and F. η 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 represented by Q, E, η, 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, η represents taste or income shocks that affect the marginal value of footprint, and F denotes the price of fuel. Miles traveled are fixed.[5]

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

---

future fuel prices. In that case, the pre-existing distortion would likely provide an efficiency justification for reverse attribute-basing, since doing so would partially mitigate the output distortion.

[5]An attribute-based fuel economy standard is properly modeled as an emissions standard that is both output-based (since total emissions can vary with miles driven) and attribute-based. Fully modeling the dependence of an emissions standard on both output and an attribute simultaneously is beyond the scope of this paper. My main analysis therefore considers output and attribute-basing one at a time rather than jointly (and therefore treats attribute-based fuel economy standards and fuel economy standards with endogenous miles traveled as separate applications) Appendix A.6 does, however, extend the results presented in section 3.1 to consider the welfare implications of attribute-basing a standard that is already output-based.

3. **Electricity generation:** Q denotes total electricity consumption, E denotes total CO_2 emissions, η 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 an aggregate, private net benefits function $B(Q, E, \eta, F)$ that captures the difference between consumers' utility (not including any externalities) and firms' production cost.[6] I denote the marginal net benefits of output Q and emissions E by the partial derivatives $B_Q(Q, E, \eta, F)$ and $B_E(Q, E, \eta, F)$.

I assume that there are diminishing net returns to both Q and E so that $B_Q(Q, E, \eta, F)$ and $B_E(Q, E, \eta, F)$ are positive at low values of Q and E but negative for large Q and E. That is, I assume that the second derivatives $B_{QQ}(Q, E, \eta, F)$ and $B_{EE}(Q, E, \eta, F)$ are both strictly negative. For instance, 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 derivative $B_{QE}(Q, E, \eta, F) \geq 0$ to reflect the fact that in most applications, reducing emissions is more costly for large Q than for small Q (starting from the same initial emissions level E).

η should be thought of as a scalar income or preference shock that affects demand for output Q, though it could also represent a marginal production cost shifter. I sign η as a "positive" shock, so that the cross-partial derivative $B_{Q\eta}(Q, E, \eta, F) > 0$. η does not affect the marginal value of emissions E (conditional on Q), however, so the cross-partial $B_{E\eta}(Q, E, \eta, F) = 0$. F should be thought of as the price (= marginal cost) of carbon-

---

[6] I show in appendix A.2 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 (as is the case in the United States for renewable portfolio standards, fuel economy standards, renewable fuel standards, and the California Low Carbon Fuel Standard); and (3) equal and constant welfare weights and marginal utility of income across all agents (per Kaplow (2012), this last condition can be relaxed if welfare weights vary only according to wage-earning ability and agents' utility functions are weakly separable in labor effort, in which case non-linear income taxes can be used to redistribute instead). Ito and Sallee (2018) shows that if this last condition is violated, then redistribution may provide a motivation for attribute-basing. I return to the question of distributional motivations in this paper's conclusion.

intensive fuel, where increased fuel prices decrease the net marginal benefit from emissions, so that $B_{EF}(Q, E, \eta, F) < 0$. F could also, in principle, represent the state of firms' emissions control technology. Conditional on E, changes in F do not affect the marginal value of Q, so that $B_{QF}(Q, E, \eta, F) = 0$.[7]

In the absence of regulation, agents in equilibrium will choose Q and E, given η and F, so that the private first-order conditions (FOCs) are satisfied: $B_Q(Q, E, \eta, F) = 0$ and $B_E(Q, E, \eta, F) = 0$.[8] As shown in appendix A.1, the implicit function theorem then yields an intuitive set of comparative statics for how the equilibrium Q and E are affected by shocks to η and F: $dQ/d\eta > 0$, $dQ/dF \leq 0$, $dE/d\eta \geq 0$, and $dE/dF < 0$ (with all inequalities strict iff $B_{QE} > 0$).

Finally, I follow Weitzman (1974) by assuming that $B(Q, E, \eta, F)$ can be well-approximated by a second-order Taylor expansion. This assumption substantially enhances the analytic tractability of the model (for instance, the comparative static derivatives above all become constants), and I maintain it throughout the paper, henceforth writing all second derivatives as, for example, B_{QQ} rather than $B_{QQ}(Q, E, \eta, F)$.

2.2 The externality and regulatory responses

The total external cost of emissions is given by ϕE, so that marginal damage is equal to ϕ, a constant that is proportional to the social cost of carbon.[9] The full-information social

---

[7]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. This 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). It is likely violated, however, for the electricity sector, since increases in the cost of carbon-intensive fuel increase the marginal cost of generation (even holding emissions fixed). Thus, the electricity application discussed in section 4.3 must account for the fact that some of the variation in F also manifests effectively as variation in η, as discussed in more detail in appendix B.3.

[8]The second order condition (SOC) sufficient for an equilibrium is that $B_{QQ}B_{EE} - B_{QE}^2 > 0$.

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

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 $B_E(Q, E, \eta, F) = \phi$. Intuitively, the planner's solution therefore 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 η and F will be in the second stage. The regulator does, however, have rational beliefs about what the distributions of η and F will be in stage 2. Then in stage 2 agents choose Q and E given the policy and the realized η 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 η and F. Either way, the underlying assumption is that the regulator cannot immediately adjust the emissions policy in response to realized shocks.[10] This assumption is consistent with U.S. emissions regulation in practice, where many years typically elapse between policy updates. For instance, the 2012 fuel economy standard revisions were locked-in until at least 2021, and the Clean Power Plan prescribed emission standards through 2030, despite the potential for large economic shocks during the intervening years.

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

It is easy to show that a Pigouvian tax τ on emissions E achieves the full-information first-best, since in that case the agents' welfare maximization problem is identical to that of

---

[10]Nor can the regulator retroactively update the policy, as in Pizer and Prest (forthcoming).

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 output-based standards that are simple affine functions of output, so that $\mu(Q) = \mu_0 + \gamma Q$. This functional form encompasses "flat" standards whenever $\gamma = 0$ and intensity standards whenever $\mu_0 = 0$ and $\gamma > 0$.

What are the welfare-maximizing levels of μ_0 and γ? 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 η. I show that the welfare-maximizing value of γ is generically non-zero and will be strictly greater than zero unless η and F have a large positive correlation.

3.1 Optimal standards under uncertainty in F

Suppose that η 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 η in the notation for the remainder of this subsection). The regulator's problem is then to choose the standard's intercept and slope parameters μ_0 and γ to maximize expected welfare:[11]

$$\max_{\mu_0, \gamma} \int_{F_L}^{F_H} \left(B(Q(\mu_0, \gamma, F), E(\mu_0, \gamma, F), F) - \phi E(\mu_0, \gamma, F) \right) w(F) dF, \tag{2}$$

$w(F)$ denotes the probability density function of F, with support on $[F_L, F_H]$.

What are the optimal parameters μ_0 and γ? It is useful to begin by building intuition for agents' choices of Q and E in the absence of regulation. The black line in figure 1, panel (a) traces out how agents' choices of Q and E decrease as F increases from F_L to F_H. Per

---

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

the comparative statics in appendix A.1, the locus of points $\{(Q(F), E(F))\}$ is a straight line in Q, E space, with a slope of $-B_{QQ}/B_{QE}$.

Panel (a) of figure 1 also depicts the socially optimal points $\{(Q^*(F), E^*(F))\}$. At the social optimum, agents internalize the externality ϕ. The effects of doing so on Q and E are equivalent to an increase in F, so for any realized F the socially optimal $(Q^*(F), E^*(F))$ is simply a southwestward shift—following along the original locus of choices—from the privately optimal $(Q(F), E(F))$. One way to then think about the planner's problem, equation (2), is that the goal is to choose values of μ_0 and γ that minimize the distance, averaged over possible realizations of F, between the socially optimal $\{(Q^*(F), E^*(F))\}$ and the choices $\{(Q(\mu_0, \eta, F), E(\mu_0, \eta, F))\}$ induced by the policy.

Figure 1, panel (b) imposes a "flat" non-output-based standard, for which the slope γ equals zero and the intercept μ_0 is set to the value μ^* that maximizes expression (2), conditional on $\gamma = 0$. Let $\hat{F}(\mu^*)$ denote the highest fuel price at which this standard binds (i.e., for $F > \hat{F}(\mu^*)$, agents will voluntarily choose to emit $E < \mu^*$). Kellogg (2018) shows that at the optimal μ^*, the expected marginal abatement cost, conditional on the standard binding, equals marginal external damage. That is, μ^* satisfies:

$$\frac{1}{W(\hat{F}(\mu^*))} \int_{F_L}^{\hat{F}(\mu^*)} B_E(Q(\mu^*, F), \mu^*, F) w(F) dF = \phi, \tag{3}$$

where $W(F)$ is the cumulative distribution function of F. The integrand in (3) represents the marginal abatement cost of meeting the binding standard, given a realization $F \leq \hat{F}$.

Note that when the flat μ^* standard binds, agents' choice of Q is always given by Q^*, as shown in figure 1, panel (b), regardless of the realization of F.[12] 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^*, μ^*) for any $F \leq \hat{F}$.

Panel (b) of figure 1 also illustrates why the emissions standard, even if set optimally, falls short of the social optimum. If the realized F is sufficiently large that the standard

---

[12]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 $\hat{F}(\mu_0, \gamma)$. The planner's problem is to choose an emissions regulation that minimizes the distance, averaged over possible realizations of F, between the socially optimal $\{(Q^*(F), E^*(F))\}$ and the choices induced by the policy. See text for details.

doesn't bind or just barely binds, the resulting emissions E (and output Q) will be greater than at the social optimum. If the realized F is low, emissions will be fixed at μ_0 even though

the social optimum calls for a greater emissions rate given the high marginal abatement cost.

Panels (c) and (d) of figure 1 illustrate agents' choices under an output-based standard that rotates the flat standard around the point (Q^*, μ^*). When $\gamma > 0$, choices of Q and E now vary with F when the standard binds, since increasing Q allows agents to increase E. The extent to which Q varies with F is proportional to γ (see appendix A.1 for formal comparative statics). The extent to which E varies with F is then proportional to γ^2, since when agents are constrained a change in E (induced by a change in F) is equal to the change in Q multiplied by γ. Overall then, shocks to F induce little variation in Q and E when γ is small (panel (c) of figure 1) but substantial variation when γ is large (panel (d)).

The fact that E can vary with F when $\gamma > 0$ creates a flexibility benefit because emissions are allowed to increase—moving closer to the social optimum—when fuel prices are low. This benefit comes at the cost, however, of distorting Q: output is too high when the standard binds. This distortion creates a standard Harberger deadweight loss triangle, the size of which is proportional to γ^2 (Ito and Sallee 2018). Thus, both the flexibility benefit and distortionary cost of output-basing are second-order in γ, the slope of the standard. The distortionary effect dominates, since the flexibility itself arises only because of the distortion to Q. The optimal standard is therefore a corner solution: flat, with $\gamma = 0$.

Appendix A.3 formally proves that the optimal standard, under uncertainty in F only, has $\gamma = 0$. This appendix also shows that, conditional on setting $\gamma > 0$, the optimal intercept μ_0 should vary so that the standard's "pivot point" is fixed at (Q^*, μ^*), as shown in panels (c) and (d) of figure 1.[13]

---

[13]The intuition relates to that driving equation (3): the optimal level of the standard should balance marginal damage with marginal abatement cost, conditional on the standard binding. Output-basing increases marginal abatement cost (accounting for the output distortion), but it also increases the sensitivity of emissions (and therefore damages) to the level of the standard. These two effects cancel each other out so that, regardless of the value of γ, the fuel price $\hat{F}$ at which the optimal standard binds does not change.

3.2 Optimal standards under uncertainty in η

I now consider optimal regulation under uncertainty in the marginal value of Q, denoted by η. This subsection assumes that F is fixed and known by the regulator; section 3.3 will consider uncertainty in both η and F. The regulator's problem is now given by:

$$\max_{\mu_0, \gamma} \int_{\eta_L}^{\eta_H} \left(B(Q(\mu_0, \gamma, \eta), E(\mu_0, \gamma, \eta), \eta) - \phi E(\mu_0, \gamma, \eta) \right) v(\eta) d\eta, \tag{4}$$

$v(\eta)$ denotes the probability density function of η, with support on $[\eta_L, \eta_H]$.

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 in appendix A.1). These choices are illustrated in figure 2, panel (a). Both Q and E increase with η, 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; 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 ϕ was isomorphic to an increase in F. This relationship does not hold for η, so that the line denoting socially optimal $\{(Q^*(\eta), E^*(\eta))\}$ 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, with $\mu_0 = \mu^*$ so that the marginal abatement cost, conditional on the standard binding, equals marginal damage. The FOC that governs this standard is given by equation (5) below, where $\hat{\eta}$ denotes the value of η at which the standard just binds, and $V(\eta)$ is the cumulative distribution function for η:

$$\frac{1}{1 - V(\hat{\eta}(\mu^*))} \int_{\hat{\eta}(\mu^*)}^{\eta_H} B_E(Q(\mu^*, \eta), \mu^*, \eta) v(\eta) d\eta = \phi. \tag{5}$$

When only F was uncertain, agents' choice of Q was fixed at Q^* whenever the flat

---

[14]The slope dE/dQ induced by variation in η 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.
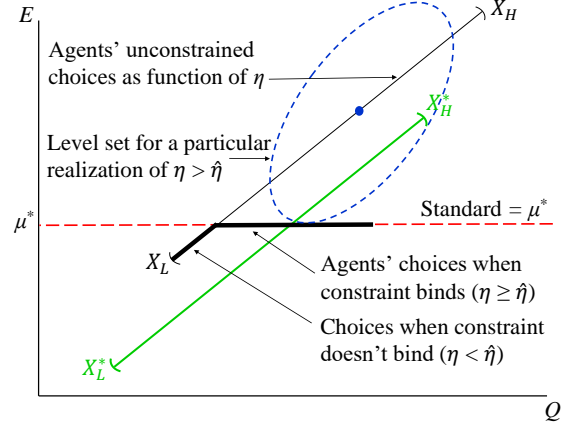
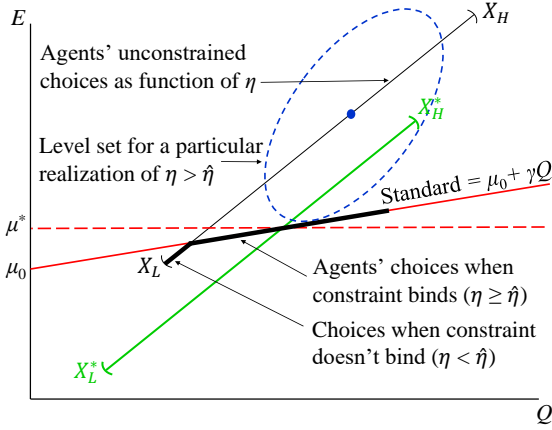**Figure 2:** Agents' choices, social optima, and regulations under uncertainty in η

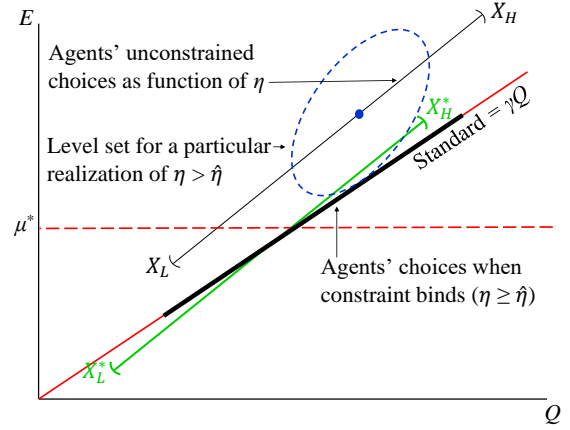**(a)** Agents' unrestricted choices and social optima
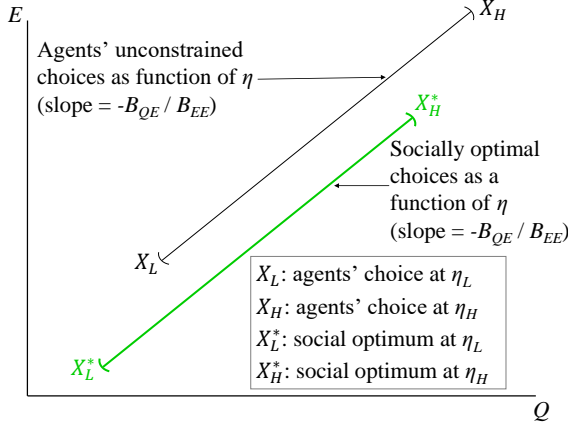
(b) Optimal flat standard

(c) Optimal 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 η 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^*. η_L and η_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 η greater than $\hat{\eta}(\mu_0, \gamma)$. The planner's problem is to choose an emissions regulation that minimizes the distance, averaged over possible realizations of η, between the socially optimal $\{(Q^*(\eta), E^*(\eta))\}$ and the choices induced by the policy. See text for details.

standard bound, regardless of the realization of F. When only η is uncertain, however, Q varies linearly with η when the standard is binding, as shown in panel (b) of figure 2 and

derived in appendix A.1. Intuitively, increases in η (the marginal value of Q) lead agents to increase Q in equilibrium, holding E fixed at μ^*.

Now consider what happens when the standard is rotated so that it is output-based, with $\gamma > 0$, as shown in figure 2, panel (c). Because Q varied linearly with η even when the standard was flat, the extent to which E varies with η is now proportional to γ rather than γ^2 (since when agents are constrained, a change in E induced by a change in η is equal to the change in Q multiplied by γ). Thus, under uncertainty in η the flexibility benefits of output-based standards are first-order in γ rather than second-order.

In contrast, the welfare loss caused by output-basing's distortion to Q remains second-order. Output-based standards—in the presence of uncertainty over the net demand for output—therefore generate first-order flexibility benefits but only second-order output distortions, implying that the optimal value of γ strictly exceeds zero. This result is proven in appendix A.4. Figure 2, panel (c) depicts an optimally-set output-based standard, which does a better job of minimizing the average distance between agents' induced choices and the social optima than does the flat standard in panel (b).[15]

This model also provides guidance on when a small versus a large amount of output-basing will be welfare-maximizing. As shown in appendix A.4, the optimal γ, which I denote γ^*, increases with:

1. The ratio of the volatility of output (induced by the variance of η) to the externality ϕ. Intuitively, the flexibility benefits of output-basing are most valuable when the uncertainty faced by the regulator is economically large.

2. The rate at which unconstrained agents vary E relative to Q in response to shocks to η. When this slope (given by $-B_{QE}/B_{EE}$) is large, output shocks can generate large changes in marginal abatement cost under a flat standard.

---

[15]The optimal output-based standard intersects agents' unconstrained choice line at a lower point than does the flat standard. This intersection is governed by FOC_{μ_0} (equation (56) in appendix A.4), which requires that $\hat{\eta}$ decrease as γ increases.

3. The insensitivity of agents' choice of Q to changes in the cost of Q (conditional on the variance of output). Intuitively, an inelastic net demand for Q implies small welfare losses from the distortion to Q induced by output-based regulation.

Finally, it is possible that γ^* 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 visually yields choices of Q and E that are closer to the first-best than does a flat standard.

3.3 Uncertainty in both η and F

Finally, consider optimal output-basing when the regulator faces uncertainty in both η and F. I show in appendix A.5 that the optimal regulatory slope γ^* then depends on the correlation between η and F. When η and F are uncorrelated, the results discussed in section 3.2 above continue to hold, and $\gamma^* > 0$. If the covariance between η and F is positive and sufficiently large relative to the variance of η, however, the optimal value of γ may be zero or even negative. Moving in the other direction, a negative correlation between η and F implies an increase in γ^* relative to the case when they are uncorrelated.

The intuition behind these results flows from the fact that when F is correlated with η, 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. When η and F are negatively correlated, positive shocks to η 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 η 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.

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

Application	Q	E	η	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.

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 vehicle attributes fixed); and (3) an output-based CO_2 standard for electric generation. Table 1 summarizes the interpretation of the variables Q, E, η, 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 η (the marginal value of output) and F (fuel prices), even though volatility in F 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.[16] My baseline simulations assume that the correlation between η 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.[17] In practice, I am not aware of any emissions standard that is indexed directly to fuel prices; thus, it seems appropriate to eliminate this motivation for output-based standards here. Nonetheless, to shed light on the effects of correlation between η and F, I include a sensitivity analysis in the electricity sector application in section 4.3.

In each of the three applications, quantifying welfare effects requires estimates of the parameters governing $B(Q, E, \eta, F)$, the externality ϕ, and the uncertainty in η and F faced by the regulator. The three subsections below—one for each application—summarize how I calibrate these parameters using estimates from previous studies and my own estimates of the historic volatility of attributes, output, and fuel prices. They then present simulated welfare

---

[16]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 γ are qualitatively unchanged from those presented below. The optimal slopes are 0.003 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.

[17]When I estimate the correlation between η and F using the relevant data for each application, I obtain correlation coefficients of -0.05, 0.45, and -0.06 for the footprint, miles traveled, and electricity applications, respectively. In each case, incorporating these correlations into the simulation does not affect the rank ordering of the flat vs. intensity standards.

outcomes from different emissions policies. Additional detail on the parameter calibration is presented in appendix B.

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 η and F.

I calibrate the model from the perspective of a regulator setting an emissions standard in 2012 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 summarize my calibration here and present additional detail in appendix B.1.

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 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. And I derive B_{QE} from the fact that the actual slope parameter γ for U.S. fuel economy regulations was drawn to match $-B_{QE}/B_{EE}$, the rate at which private agents increase E with Q.[18]

B_{EF} in this application represents the (negative) lifetime discounted miles traveled for

---

[18]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 $-B_{QE}/B_{EE}$.

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
σ_F	$0.35 / gal	historic volatility of gasoline prices
σ_η	$41.70 / ft^2	historic volatility of vehicle footprints
ϕ	$486 / (gal/100mi)	Kellogg (2018) externality of $0.42/gal, converted to vehicle lifetime using Busse *et al.* (2013)

new vehicles. I derive B_{EF} 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).

To obtain ϕ, 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 $-B_{EF}$.

I model the realized fuel price F as normally distributed, with a mean of F_0 and a standard deviation σ_F. I set F_0 equal to the average 2012 retail gasoline price of $3.69/gallon (real $2012).[19] To calibrate σ_F, I estimate, over the ten-year regulatory horizon, the volatility of the three-year moving average of historic monthly U.S. gasoline prices.[20] 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. I set σ_F equal to the average volatility across all horizons $t \in [1, 120]$, yielding $\sigma_F = 0.35/gal.[21]

---

[19] 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).

[20] 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. Were I to instead use the raw monthly data, I would estimate $\sigma_F = 0.65/gal, 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.

[21] 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

I normalize the expected value of η to zero and calibrate σ_η using the fact that under a flat fuel economy standard, σ_η is equal to $B_{QQ}/B_{Q\eta}$ times the standard deviation of Q, which I denote σ_Q (see equation (18) in appendix A.1). I adopt the normalization $B_{Q\eta} = 1$ and obtain an estimate of σ_Q 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 σ_Q by taking long differences, as I did for σ_F. Instead, I estimate an AR1 process for Q, use the resulting estimates to project annual volatility in Q out over a ten-year horizon, and then obtain σ_Q by averaging volatility over the ten years.

Given these inputs, figure 3 presents footprint, fuel economy, and welfare outcomes under various forms of fuel economy regulation. The "x" denotes $\bar{Q}$ and $\bar{E}$: the baseline 2012 footprint and fuel economy (56.8 ft^2 and 4.37 gal/100mi, respectively) that would have been chosen in the absence of regulation.[22] The η 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 η and F, respectively. The representation of agents' choices of Q and E when unconstrained by regulation takes the form of an ellipse rather than a line (as in figures 1 and 2), owing to the presence of uncertainty in both η and F. The area encircled by the solid black ellipse denotes agents' unconstrained choices given realizations of η 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, dotted black line (which is mostly hidden behind the solid 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 optimum. This shortfall is a consequence of the large uncertainty in F, which results in substantial expected variation in marginal abatement costs when the standard binds.
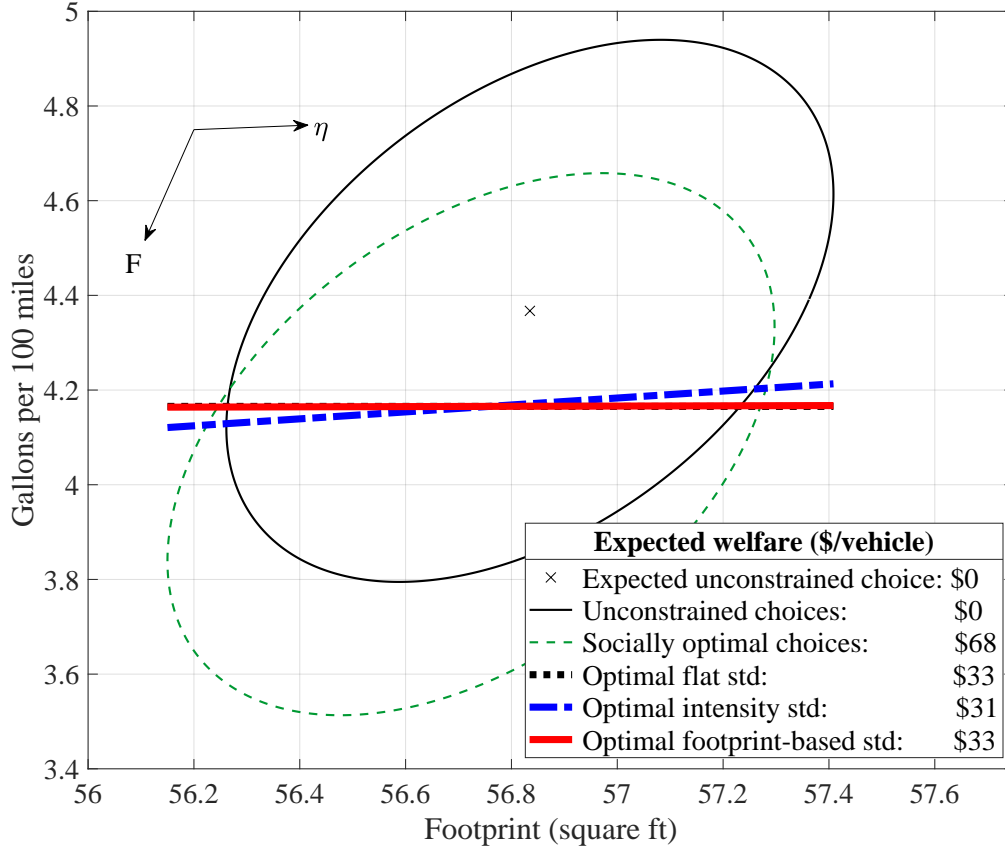
---

implement and because it more tightly connects to the model presented in sections 2 and 3.

[22]Appendix B.1 discusses how I obtain values for $\bar{Q}$ and $\bar{E}$.

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

Expected welfare ($/vehicle)	
× Expected unconstrained choice:	$0
— Unconstrained choices:	$0
– – Socially optimal choices:	$68
▪▪▪ Optimal flat std:	$33
▪–▪ Optimal intensity std:	$31
— Optimal footprint-based std:	$33

Note: Expected welfare is normalized to zero for agents' unconstrained choices. The solid black and dashed green ellipses indicate agents' privately optimal and socially optimal (respectively) choices of footprints and fuel economy associated with realizations of η and F that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The η and F arrows indicate the changes in footprint and fuel economy induced by a one standard deviation change in η and F when agents are unconstrained. The bold straight lines denote the optimal flat standard (dotted black), optimal intensity standard (dash-dot blue), and optimal footprint-based standard (solid red). See text for details.

The optimal amount of attribute-basing is small: the optimal value of γ is 0.003 gallons per 100 miles per ft^2, equal to only 6% 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 dash-dot 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 η, 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). Even with a near-doubling of σ_η to \$76.83/ft^2, the optimal attribute slope γ is still only 0.008 gallons per 100 miles per ft^2.[23]

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 η.[24] I summarize the calibration here and present details in appendix B.2.

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, $B(Q, E, \eta, F) =$

---

[23]$\sigma_\eta = $ \$76.83/ft^2 is obtained by assuming footprints follow a random walk with drift, as discussed in appendix B.1, rather than the AR1 process used in the main estimates.

[24]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 (η). Realized utility for each vehicle will be a function of realized shocks.

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 (2020)
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
σ_F	\$0.35 / gal	historic volatility of gasoline prices
σ_η	\$55.50 / 1000mi	historic volatility of miles traveled
ϕ	\$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.

$U(Q, \eta) - EF - C(E/Q)$.[25] Taking derivatives then yields:[26]

$$B_{QQ} = U_{QQ} - C''E^2Q^{-4} - 2C'EQ^{-3} \tag{6}$$

$$B_{QE} = C''EQ^{-3} + C'Q^{-2} \tag{7}$$

$$B_{EE} = -C''Q^{-2} \tag{8}$$

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 $dQ/dF = E/(QU_{QQ})$. I use an elasticity of -0.081 from Gillingham's (2020) recent survey of the literature on the rebound effect.[27]

C'' is the same object as B_{EE} from the footprint application in table 2. I calculate C'

---

[25]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.

[26]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 $Q_0 = 115644$ 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).

[27]Of the papers on the rebound effect that Gillingham (2020) surveys, -0.081 is the average elasticity among papers that use odometer readings as the primary data source. The average elasticity across all papers that Gillingham (2020) surveys is -0.141. Using this elasticity rather than the baseline elasticity of -0.081 only modestly changes the results: the optimal regulation slope γ is 0.73 gallons per 100 miles rather than 0.74 gallons per 100 miles.

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 $F_0 = \$3.69/\text{gallon}$. 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 $\phi = \$0.42/\text{gallon}$.

The calibration of σ_F is the same as in the footprint-based standard application in section 4.1. To calibrate σ_η, I use the fact that σ_η is given by the volatility in miles traveled, σ_Q, divided by $dQ/d\eta$.[28] To compute σ_Q, I obtain historical data on average U.S. vehicle miles traveled from the Federal Reserve Bank of St. Louis. I adjust the miles traveled time series by subtracting off the effects of shocks to fuel prices F, in order to isolate volatility induced by η. I then compute σ_Q by extrapolating annual volatility in adjusted miles traveled over a ten-year horizon, under an assumption that miles traveled evolves as a random walk with drift.

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 γ of 0.74 gallons per 100 miles is considerably less than the slope of 4.06 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 η shocks, since consumers choose more efficient

---

[28]$dQ/d\eta$ for constrained agents is given in appendix A.1, equation (18), where γ_0 denotes the 2012 average fleet-wide fuel economy of 4.01 gallons per 100 miles.

Figure 4: Gasoline consumption, miles traveled, and welfare under fuel economy standards with endogenous miles traveled
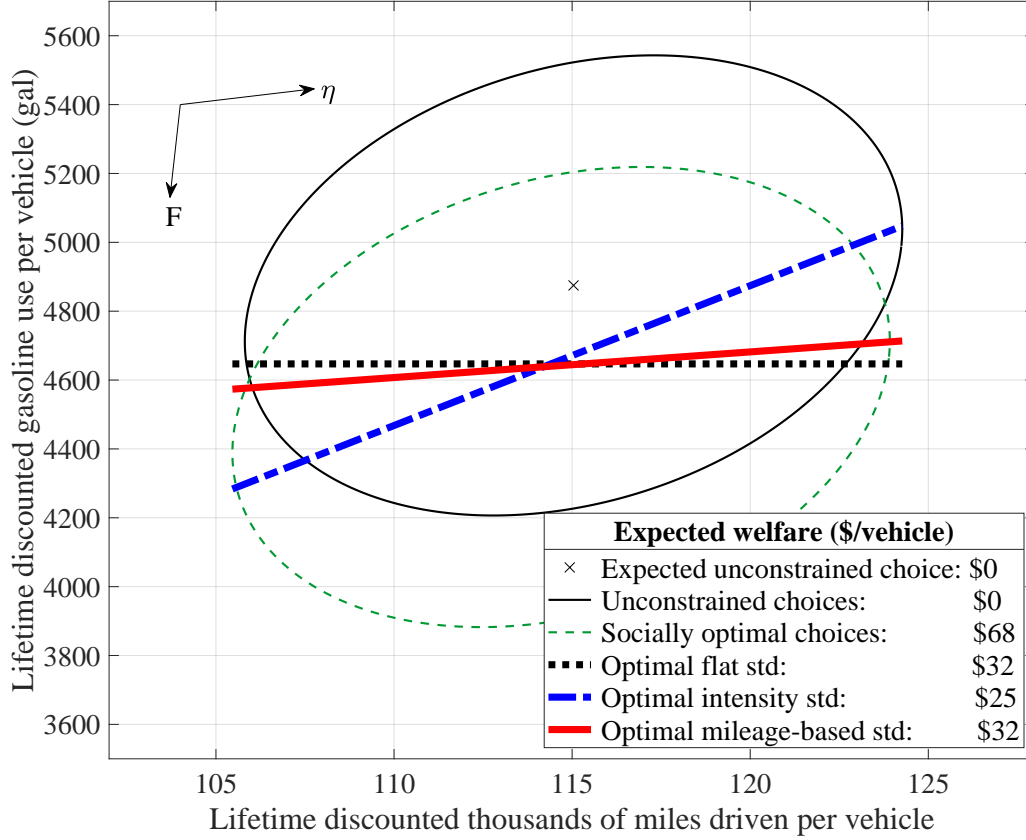
Note: Expected welfare is normalized to zero for agents' unconstrained choices. The solid black and dashed green ellipses indicate agents' privately optimal and socially optimal (respectively) choices of miles traveled and fuel use associated with realizations of η and F that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The η and F arrows indicate the changes in miles traveled and fuel use induced by a one standard deviation change in η and F when agents are unconstrained. The bold straight lines denote the optimal flat standard (dotted black), optimal intensity standard (dash-dot blue), and optimal output-based standard (solid red). See text for details.

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 gasoline prices 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 η, representing the marginal value of electric power, and $F = -P_g$, representing the negative price of natural gas. I use the natural gas price to capture F 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).[29] $B(Q, E, \eta, F)$ 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.[30]

η 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.* (2019) 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, and I discuss my calibration procedure in greater detail in appendix B.3. To calibrate B_{QQ}, I use the estimated demand elasticity of -0.09 from Ito (2014), which uses residential billing data from California and incorporates responses to price lags of up to four months.[31] As a sensitivity analysis, I also examine results that use a long-run demand elasticity of -0.27 from Deryugina, MacKay, and Reif (2020).[32]

---

[29]I set F equal to *negative* P_g because low natural gas prices cause substitution from coal to gas, which reduces (rather than increases) CO_2 emissions holding Q fixed.

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

[31]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, in which case I would underestimate the total magnitude of B_{QQ}.

[32]Deryugina, MacKay, and Reif's (2020) elasticity of -0.27 is a two-year elasticity. That paper's six-month elasticity matches the estimate of -0.09 from Ito (2014).

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

Parameter	Value	Main sources
B_{QQ}	-\$344702 / 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
σ_F	\$2.59 / mmBtu	historic volatility of natural gas prices
σ_η	\$55.4 million / TWh	historic volatility of electricity consumption
ϕ	\$38 million / Mmton	Interagency Working Group (2013)

To calibrate B_{EE}, I use Cullen and Mansur's (2017) estimate that a \$40 per ton tax on CO_2 will reduce emissions by 7.9%, holding electricity output fixed. To obtain B_{QE}, I use the relationship $dE/dQ = -B_{QE}/B_{EE}$, and I obtain dE/dQ by leveraging 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.[33] 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.

To calculate σ_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. I calculate σ_F from these data using the same long-difference procedure discussed in section 4.1 to calculate σ_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.[34] I estimate $\sigma_F = \$2.59/\text{mmBtu}$.

I estimate σ_η using annual data on net electricity generation from the electric power industry, dating back to 1990, available from the EIA. The procedure I use is similar to that from the miles traveled application in section 4.2. First, I adjust the generation time series for changes induced by fluctuations in the price of natural gas. I then compute σ_Q

---

[33]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.

[34]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.

by extrapolating annual volatility in adjusted generation over a 15-year horizon, under an assumption that generation evolves as a random walk with drift. I convert σ_Q to σ_η using the fact that the volatility of Q equals $dQ/d\eta$ times σ_η.[35]

The model's output is presented in figure 5. For the U.S. electricity market, I find that an optimally-set intensity standard modestly outperforms an optimal flat standard, yielding expected welfare gains of \$1.42 rather than \$1.24 billion/year (relative to no regulation). The optimal output-based regulation has a slope parameter $\gamma = 0.32$ 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, stems primarily from the fact that the variance in η (the marginal value of electricity) is relatively large in this application. The variation in electricity use Q induced by σ_η is such that the standard deviation of Q, absent regulation, is 4.9% of its expected value.[36]

Together, the uncertainty in η and F is sufficiently large relative to the externality ϕ that the optimal flat, intensity, and output-based standards all would bind only slightly more than 50% of the time. These results are related to those presented in Borenstein *et al.* (2019), 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. In addition, the substantial volatility of both η 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 \$3.28 billion/year.

---

[35] $dQ/d\eta$ is given by $-B_{EE}B_{Q\eta}/(B_{QQ}B_{EE} - B_{QE}^2)$, as derived for unconstrained agents in appendix A.1, equation (12).

[36] Relative to Borenstein *et al*'s (2019) study of the California electricity market, this uncertainty is small. Borenstein *et al.* (2019) forecasts non-hydro electricity generation for 2020, starting from a base year of 2012, and finds that the standard deviation of the forecast is 17% of its expected value (see column (1) of table 2 in that paper). Since shocks to generation are not perfectly correlated across states, it is not surprising that forecast uncertainty for a single state is appreciably larger than uncertainty over the entire United States. Borenstein *et al*'s (2019) estimate also incorporates parameter uncertainty, whereas the estimate I compute assumes that generation evolves as a random walk.

Figure 5: CO_2 emissions, electricity consumption, and welfare under output-based electricity emission standards
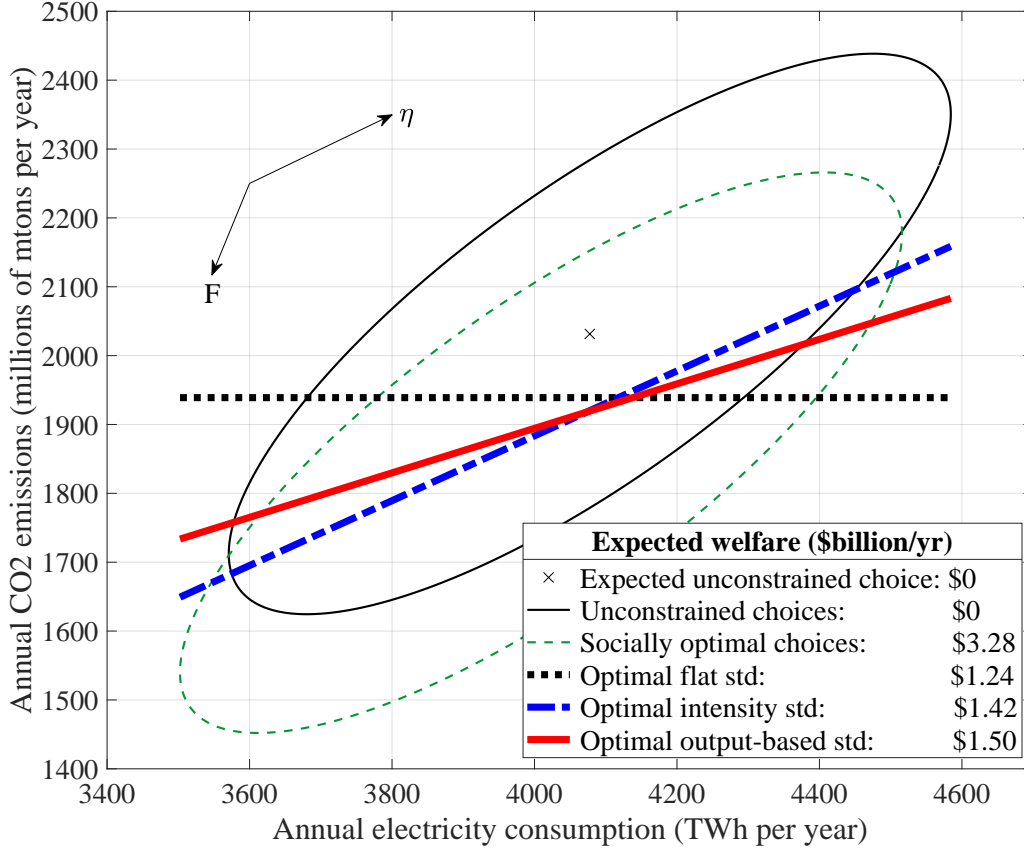
Note: Expected welfare is normalized to zero for agents' unconstrained choices. The solid black and dashed green ellipses indicate agents' privately optimal and socially optimal (respectively) choices of electricity consumption and CO_2 emissions associated with realizations of η and F that lie on the iso-pdf of their bivariate normal distribution that encompasses 95% of the probability mass. The η and F arrows indicate the changes in electricity consumption and CO_2 emissions induced by a one standard deviation change in η and F when agents are unconstrained. The bold straight lines denote the optimal flat standard (dotted black), optimal intensity standard (dash-dot blue), and optimal output-based standard (solid red). See text for details.

The welfare ranking of the flat versus intensity standard is sensitive to parameter inputs. In particular, changes to demand uncertainty σ_η, the electricity demand elasticity, or the correlation ρ between demand shocks η and fuel price shocks F can change the policy welfare ordering:

- **Demand uncertainty σ_η.** If I halve σ_η and re-simulate the model, the optimal slope

γ falls by roughly half, to 0.15 Mmtons CO_2 per TWh, and a flat standard outperforms the intensity standard. Conversely, if I double σ_η the optimal gamma increases to 0.44 Mmtons CO_2 per TWh, and the flat standard achieves less than half the expected welfare gain captured by the intensity standard.[37]

- **Demand elasticity**. If I use the long-run elasticity from Deryugina, MacKay, and Reif (2020) of -0.27 rather than the elasticity of -0.09 from Ito (2014), the flat standard achieves an expected welfare gain of $1.94 billion per year, relative to $1.61 billion per year under the intensity standard.[38] Still, even in this case I find that the welfare-maximizing standard is substantially output-based, with a slope of 0.26 Mmtons CO_2 per TWh.

- **Correlation between η and F**. The baseline simulation assumes that the η and F shocks are uncorrelated. If I instead impose a modest correlation of 0.25, that is sufficient to reverse the welfare ordering: the expected welfare outcomes from the optimal flat and intensity standards are $1.54 and $1.44 billion/year, respectively.

These sensitivities highlight the potential value of future work that could more faithfully estimate a model of the U.S. electricity sector and then simulate the impacts of flat or output-based standards. For instance, the VAR-based forecasting strategy used in Borenstein *et al.* (2019), if applied to the entire United States, could provide better estimates of both future demand uncertainty and its correlation with fuel price uncertainty (for potentially multiple fuels, not just natural gas). A model of dynamic demand adjustment could bridge the short and long-run demand elasticities, rather than just use one or the other as I do here. Finally, using detailed data on individual generators' fuel use and marginal costs could improve the

---

[37]When σ_η is halved relative to the value in table 4, the expected welfare from the flat, intensity, and optimal output-based standards are $1.58, $1.42, and $1.62 billion/year, respectively. Conversely, when σ_η is doubled, the welfare from the flat, intensity, and optimal output-based standards are $0.52, $1.41, and $1.42 billion/year, respectively.

[38]With a higher demand elasticity, all welfare outcomes increase. The Pigouvian tax achieves a gain of $6.46 billion per year.

fidelity of the supply side of the model and avoid the need to assume constant values for B_{QE}, B_{EE}, and B_{EF}.

5   Conclusions

The consensus in the economics 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 depends on the underlying economic environment and especially on the uncertainty of future demand. For U.S. attribute-based fuel economy standards, I find that the welfare-maximizing amount of attribute-basing is negligible, owing primarily to the limited uncertainty regarding the demand for vehicle size. 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 could more precisely model sector-specific details such as generation unit-specific dispatch, short vs. long-run demand elasticities, and future uncertainty in the demand for power. Future research could also extend this paper's findings to other output or attribute-based policies such as renewable portfolio standards or appliance energy efficiency standards,

evaluating in each case the relevant uncertainties and other key parameters.

Notwithstanding the above results, 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.

A natural path for future work would then be to link the normative results in this paper to the positive literature in political economy on why particular environmental policies are or are not adopted. Recent reviews of this literature by Keohane, Revesz, and Stavins (1998), Oates and Portney (2003), and Schmalensee and Stavins (2019) emphasize how cap-and-trade schemes that grandfather permits are more likely to build durable pro-regulation constituencies than are emissions taxes, despite the efficiency benefits of the latter. Rabe (2018) additionally points out that cap-and-trade may be politically advantaged because its costs are hidden relative to the explicit prices imposed via a tax. This literature has had less to say, however, on the prevalence of intensity, output-based, and attribute-based regulations, especially relative to cap-and-trade programs with targeted grandfathering or rebates. The present paper speaks to possible efficiency justifications for such policies in the presence of uncertainty. Another potentially fruitful line of research would be to analyze the distributional politics of output-based standards, perhaps linking to the broader political economy literature on distortionary distribution mechanisms (Grossman and Helpman 1994, Coate and Morris 1995, Acemoglu and Robinson 2001, Glaeser and Ponzetto 2014).

References

[1] Acemoglu, Daron and James A. Robinson (2001), "Inefficient Redistribution", *American Political Science Review* 95, 649-661.

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

[3] 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.

[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 (2019), "Expecting the Unexpected: Emissions Uncertainty and Environmental Market Design", *American Economic Review* 109(11), 3953-3977.

[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] Coate, Stephen and Stephen Morris (1995), "On the Form of Transfers to Special Interests", *Journal of Political Economy* 6, 1210-1235.

[9] 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.

[10] Deryugina, Tatyana, Alexander MacKay, and Julian Reif (2020), "The Long-Run Dynamics of Electricity Demand: Evidence from Municipal Aggregation", *American Economic Journal: Applied Economics* 12(1), 86-114.

[11] 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.

[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 (2020), "The Rebound Effect and the Rollback of Fuel Economy Standards", *Review of Environmental Economics and Policy* 14(1), 136-142.

[14] Glaeser, Edward L. and Giacomo A.M. Ponzetto (2014), "Shrouded Costs of Government: The Political Economy of State and Local Public Pensions", *Journal of Public Economics* 116, 89-105.

[15] 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.

[16] Grossman, Gene M. and Elhanan Helpman (1994), "Protection for Sale", *American Economic Review* 84, 833-850.

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

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] Kaplow, Louis (2012), "Optimal Control of Externalities in the Presence of Income Taxation", *International Economic Review* 53, 487-509.

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

[25] Keohane, Nathaniel O., Richard L. Revesz, and Robert N. Stavins (1998), "The Choice of Regulatory Instruments in Environmental Policy", *Harvard Environmental Law Review* 22, 313-368.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

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

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

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

[33] Oates, Wallace E. and Paul R. Portney (2003), "The Political Economy of Environmental Policy", in *Handbook of Environmental Economics* 1, Karl-Goran Mäler and Jeffrey Vincent, Eds. North Holland.

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

[35] Pizer, William A. and Brian Prest (forthcoming), "Prices versus Quantities with Policy Updating", *Journal of the Association of Environmental and Resource Economists.*

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

[37] Rabe, Barry G. (2018), *Can We Price Carbon?* Cambridge, MA: MIT Press.

[38] 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.

[39] Schmalensee, Richard and Robert N. Stavins (2019), "Policy Evolution under the Clean Air Act", *Journal of Economic Perspectives* 33, 27-50.

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

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

Online appendix

A Theoretical model

A.1 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:

$$\max_{Q,E} B(Q, E, \eta, F). \tag{9}$$

The FOCs for this problem are given by:

$$\text{FOC}_Q : B_Q(Q, E, \eta, F) = 0 \tag{10}$$
$$\text{FOC}_E : B_E(Q, E, \eta, F) = 0. \tag{11}$$

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:

$$\frac{dQ}{d\eta} = \frac{-B_{EE}B_{Q\eta}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{12}$$

$$\frac{dQ}{dF} = \frac{B_{QE}B_{EF}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{13}$$

$$\frac{dE}{d\eta} = \frac{B_{QE}B_{Q\eta}}{B_{QQ}B_{EE} - B_{QE}^2} \tag{14}$$

$$\frac{dE}{dF} = \frac{-B_{QQ}B_{EF}}{B_{QQ}B_{EE} - B_{QE}^2}. \tag{15}$$

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

$$\max_{Q} B(Q, \mu_0 + \gamma Q, \eta, F). \tag{16}$$

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

$$\text{FOC}_Q : \ B_Q(Q, E(Q), \eta, F) + \gamma B_E(Q, E(Q), \eta, F) = 0. \tag{17}$$

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 μ_0, γ, η, and F):

$$\frac{dQ}{d\eta} = \frac{-B_{Q\eta}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{18}$$

$$\frac{dQ}{dF} = \frac{-B_{EF}\gamma}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{19}$$

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

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

$$\frac{dQ}{d\mu_0} = \frac{-(B_{EE}\gamma + B_{QE})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{22}$$

$$\frac{dE}{d\mu_0} = 1 + \gamma\frac{dQ}{d\mu_0} = \frac{B_{QE}\gamma + B_{QQ}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{23}$$

$$\frac{dQ}{d\gamma} = \frac{-(B_E + B_{EE}Q\gamma + B_{QE}Q)}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \tag{24}$$

$$\frac{dE}{d\gamma} = Q + \gamma\frac{dQ}{d\gamma} = \frac{-B_E\gamma + B_{QE}Q\gamma + B_{QQ}Q}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}}. \tag{25}$$

A.2 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 ϕE, 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 constant and 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 η 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 γ, FOC_Q in equation (17) 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 τ, we have $B_Q^i = -\gamma\tau$ and $B_E^i = \tau \ \forall i$. Thus, given values for η and F, I can model agents' welfare under an emissions standard by instead modeling a tax τ on E and a tax $\tau_Q \equiv -\gamma\tau$ on Q.

Next, I show that imposition of a tax τ 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:

$$\text{FOC}_Q : B_Q(Q, E, \eta, F) = \tau_Q \tag{26}$$

$$\text{FOC}_E : B_E(Q, E, \eta, F) = \tau. \tag{27}$$

Following appendix A.1, application of the implicit function theorem yields the following

comparative statics:

$$\frac{dQ_i}{d\tau_Q} = \frac{B_{EE}}{B_{QQ}B_{EE} - B_{QE}^2}; \quad \frac{dQ_i}{d\tau} = \frac{-B_{QE}}{B_{QQ}B_{EE} - B_{QE}^2}; \tag{28}$$

$$\frac{dE_i}{d\tau_Q} = \frac{-B_{QE}}{B_{QQ}B_{EE} - B_{QE}^2}; \quad \frac{dE_i}{d\tau} = \frac{B_{QQ}}{B_{QQ}B_{EE} - B_{QE}^2}. \tag{29}$$

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 η and F). Using a second-order Taylor approximation around $Q_0, E_0, \bar{\eta}$, and $\bar{F}$, social welfare is then given by:[39]

$$\begin{aligned}
\int_i (B^i(Q_i, E_i, \eta, F) - \phi E_i)di = &\int_i B_{Q0}^i(Q_i - Q_0)di + \int_i B_{E0}^i(E_i - E_0)di \\
&+ \frac{1}{2}B_{QQ}\int_i (Q_i - Q_0)^2 di + \frac{1}{2}B_{EE}\int_i (E_i - E_0)^2 di \\
&+ B_{QE}\int_i (Q_i - Q_0)(E_i - E_0)di + B_{Q\eta}\int_i (Q_i - Q_0)(\eta - \bar{\eta})di \\
&+ B_{EF}\int_i (E_i - E_0)(F - \bar{F})di - \phi E,
\end{aligned} \tag{30}$$

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 (30) 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

---

[39]The assumption of equal and constant marginal utility of income (and welfare weights) across agents is necessary for equation (30) 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:

$$\int_i (B^i(Q_i, E_i, \eta, F) - \phi E_i)di = [B_{Q0}(Q - Q_0) + B_{E0}(E - E_0) + \frac{1}{2}B_{QQ}(Q - Q_0)^2$$
$$+ \frac{1}{2}B_{EE}(E - E_0)^2 + B_{QE}(Q - Q_0)(E - E_0)$$
$$+ B_{Q\eta}(Q - Q_0)(\eta - \bar{\eta}) + B_{EF}(E - E_0)(F - \bar{F}) - \phi E]$$
$$+ [\int_i B_{Q0}^i(Q_i - Q)di + \int_i B_{E0}^i(E_i - E)di$$
$$+ \frac{1}{2}B_{QQ}\int_i(Q_i - Q)^2 di + \frac{1}{2}B_{EE}\int_i(E_i - E)^2 di$$
$$+ B_{QE}\int_i(Q_i - Q)(E_i - E)di] \tag{31}$$

The first bracketed term in equation (31) 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.

A.3   Welfare-maximizing standard under uncertainty in F

This appendix proves that, when there is uncertainty only in F, the optimal standard has $\gamma = 0$, as discussed in section 3.1 in the main text. The argument closely follows appendix B.3 in Kellogg (2018).

It will be useful to begin by developing simple expressions for B_E and B_Q while the agents are constrained. Given μ_0 and γ, 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:

$$B_Q(Q, E, F) = B_{QE}(E - \hat{E}) + B_{QQ}(Q - \hat{Q}) \tag{32}$$

On the $E = \mu_0 + \gamma Q$ standard, use equations (19) and (21) to replace $E - \hat{E}$ and $Q - \hat{Q}$ in (32), yielding:

$$B_Q(Q, E, F) = \frac{-\gamma(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}}B_{EF}(F - \hat{F}). \tag{33}$$

In addition, equation (17) implies that $B_E = -B_Q/\gamma$, implying:

$$B_E(Q, E, F) = \frac{B_{QE}\gamma + B_{QQ}}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}}B_{EF}(F - \hat{F}). \tag{34}$$

We now turn to the regulator's problem given by equation (2) in the main text, which

can be expanded to:

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

$$+ \int_{\hat{F}(\mu_0, \gamma)}^{F_H} \left(B(Q(F), E(F), F) - \phi E(F) \right) w(F) dF. \tag{35}$$

The FOCs for this problem are given by (36) and (37) 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 μ_0, γ, and F:

$$\text{FOC}_{\mu_0} : \int_{F_L}^{\hat{F}(\mu_0, \gamma)} \left(B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) w(F) dF = 0; \tag{36}$$

$$\text{FOC}_{\gamma} : \int_{F_L}^{\hat{F}(\mu_0, \gamma)} \left(B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) w(F) dF = 0. \tag{37}$$

We can simplify FOC_{μ_0} using equations (17), (22), (23), and (34). Streamlining notation by defining $S \equiv B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$, we have:

$$\text{FOC}_{\mu_0} : \int_{F_L}^{\hat{F}} \left(B_E(\frac{dE}{d\mu_0} - \gamma \frac{dQ}{d\mu_0}) - \phi \frac{dE}{d\mu_0} \right) w(F) dF = 0 \tag{38}$$

$$\Leftrightarrow \frac{1}{S} \int_{F_L}^{\hat{F}} \left(\frac{(B_{QE}\gamma + B_{QQ})B_{EF}(F - \hat{F})S}{S} - \phi(B_{QE}\gamma + B_{QQ}) \right) w(F) dF = 0 \tag{39}$$

$$\Leftrightarrow \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{F_L}^{\hat{F}} \left(\phi + B_{EF}(\hat{F} - F) \right) w(F) dF = 0. \tag{40}$$

Equation (40) implies that μ_0 should be set so that the fuel price $\hat{F}$ at which the standard binds is invariant to γ. To see this result, observe that the integral in equation (40) will only equal zero at one unique value of $\hat{F}$, regardless of the value of γ.[40] The invariance of $\hat{F}$ to γ then implies that as the regulator changes γ, the optimal "pivot point" in Q, E space about which the standard rotates is the point (Q^*, μ^*) depicted in figure 1, panels (c) and (d).

Now work with the FOC for γ, given by equation (37). Applying equations (17), (24),

---

[40] Proof: for $\hat{F}$ close to F_L, the integral is positive and increasing in $\hat{F}$. As $\hat{F}$ gets larger, the $B_{EF}(\hat{F} - F)$ term dominates the ϕ term, and the integral then decreases in $\hat{F}$, eventually crossing zero at a unique value of $\hat{F}$.

(25), and (34), we obtain:

$$\text{FOC}_\gamma : \quad \int_{F_L}^{\hat{F}} \left(B_E(\frac{dE}{d\gamma} - \gamma\frac{dQ}{d\gamma}) - \phi\frac{dE}{d\gamma} \right) w(F)dF = 0 \tag{41}$$

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

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left((B_{EF}(F - \hat{F})Q - \phi Q)S + \phi B_{EF}(F - \hat{F})\gamma \right) w(F)dF = 0. \tag{42}$$

Now use $\text{FOC}_{\mu 0}$ (equation (40)) 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 (42) to:

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left((Q - \hat{Q})(B_{EF}(F - \hat{F}) - \phi)S + \phi^2\gamma \right) w(F)dF = 0. \tag{43}$$

Then apply the derivative dQ/dF (equation (19)) to eliminate the $Q - \hat{Q}$ term:

$$\Leftrightarrow \frac{B_{QE}\gamma + B_{QQ}}{S^2} \int_{F_L}^{\hat{F}} \left(-B_{EF}\gamma(F - \hat{F})(B_{EF}(F - \hat{F}) - \phi) + \phi^2\gamma \right) w(F)dF = 0. \tag{44}$$

Applying $\text{FOC}_{\mu 0}$ (equation (40)) again yields:

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

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 (45) as follows, defining σ_{Fc}^2 is the variance of F conditional on the standard binding:[41]

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

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

$$= W(\hat{F})(B_{eF}^2\sigma_{Fc}^2 + \phi^2), \tag{46}$$

where the last line makes use of equation (40) and the definition of $\dot{F}$.

Finally, substitute equation (46) into equation (45) to obtain:

$$\text{FOC}_\gamma : \quad \frac{\gamma W(\hat{F})(B_{QE}\gamma + B_{QQ})}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2}(\phi^2 - B_{EF}^2\sigma_{Fc}^2) = 0. \tag{47}$$

---

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

Equation (47) is solved by setting $\gamma = 0$ so that the optimal standard, when there is only uncertainty in F, is flat rather than output-based.[42] 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 γ.

A.4 Welfare-maximizing standard under uncertainty in η

This appendix proves that, when there is uncertainty only in η, the optimal standard has $\gamma > 0$, as discussed in section 3.2 in the main text. It also derives the comparative statics discussed at the end of that section.

It is useful to begin by developing simple expressions for B_E and B_Q while agents are constrained. Given μ_0 and γ, let $\hat{\eta}$ denote the value of η (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:

$$B_E(Q, E, \eta) = B_{QE}(Q - \hat{Q}) + B_{EE}(E - \hat{E}) \tag{48}$$

On the $E = \mu_0 + \gamma Q$ standard, use equations (18) and (20) to replace $Q - \hat{Q}$ and $E - \hat{E}$ in (48), yielding:

$$B_E(Q, E, \eta) = \frac{-(B_{QE} + B_{EE}\gamma)}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} B_{Q\eta}(\eta - \hat{\eta}). \tag{49}$$

We now turn to the regulator's problem given by equation (4) in the main text, which can be expanded to:

$$\max_{\mu_0, \gamma} \int_{\eta_L}^{\hat{\eta}(\mu_0, \gamma)} \left(B(Q(\eta), E(\eta), \eta) - \phi E(\eta) \right) v(\eta) d\eta$$

$$+ \int_{\hat{\eta}(\mu_0, \gamma)}^{\eta_H} \left(B(Q(\mu_0, \gamma, \eta), E(\mu_0, \gamma, \eta), \eta) - \phi E(\mu_0, \gamma, \eta) \right) v(\eta) d\eta. \tag{50}$$

The regulator's FOCs for μ_0 and γ are given by equations (51) and (52) (suppressing the dependence of B_Q, B_E, $dQ/d\mu_0$, $dQ/d\gamma$, $dE/d\mu_0$, and $dE/d\gamma$ on μ_0, γ, and η):

$$\text{FOC}_{\mu_0} : \int_{\hat{\eta}(\mu_0, \gamma)}^{\eta_H} \left(B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) v(\eta) d\eta = 0; \tag{51}$$

$$\text{FOC}_{\gamma} : \int_{\hat{\eta}(\mu_0, \gamma)}^{\eta_H} \left(B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) v(\eta) d\eta = 0. \tag{52}$$

We can simplify FOC_{μ_0} using equations (17), (22), (23), and (49). Streamlining notation

---

[42] $\gamma = -B_{QQ}/B_{QE}$ is also a solution to FOC_γ. Kellogg (2018) shows, however, that only the $\gamma = 0$ solution satisfies the SOC, so long as μ_0 is chosen optimally (guaranteeing that $\phi^2 > B_{EF}^2 \sigma_{Fc}^2$).

by defining $S \equiv B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$, we have:

$$\text{FOC}_{\mu_0} : \int_{\hat{\eta}}^{\eta_H} \left(B_E(\frac{dE}{d\mu_0} - \gamma\frac{dQ}{d\mu_0}) - \phi\frac{dE}{d\mu_0} \right) v(\eta)d\eta = 0 \tag{53}$$

$$\Leftrightarrow \frac{1}{S} \int_{\hat{\eta}}^{\eta_H} \left(\frac{-(B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta})S}{S} - \phi(B_{QE}\gamma + B_{QQ}) \right) v(\eta)d\eta = 0 \tag{54}$$

$$\Leftrightarrow \frac{-1}{S} \int_{\hat{\eta}}^{\eta_H} \left((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}) \right) v(\eta)d\eta = 0 \tag{55}$$

$$\Leftrightarrow \frac{-(B_{QE}\gamma + B_{QQ})}{B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}} \int_{\hat{\eta}}^{\eta_H} \left(\frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}}B_{Q\eta}(\eta - \hat{\eta}) + \phi \right) v(\eta)d\eta = 0. \tag{56}$$

Note that, because γ appears inside the integral in equation (56), the value of $\hat{\eta}$ at which the optimal standard binds will be a function of γ.

Now work with the FOC for γ, given by equation (52). Applying equations (17), (24), (25), and (49), we obtain:

$$\text{FOC}_\gamma : \int_{\hat{\eta}}^{\eta_H} \left(B_E(\frac{dE}{d\gamma} - \gamma\frac{dQ}{d\gamma}) - \phi\frac{dE}{d\gamma} \right) v(\eta)d\eta = 0 \tag{57}$$

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

$$\Leftrightarrow \frac{-1}{S^2} \int_{\hat{\eta}}^{\eta_H} (((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))QS$$
$$+ \phi\gamma(B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}))v(\eta)d\eta = 0 \tag{58}$$

Now use FOC_{μ_0} (equation (55)) 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 (58) to:

$$\Leftrightarrow \frac{-1}{S^2} \int_{\hat{\eta}}^{\eta_H} (((B_{QE} + B_{EE}\gamma)B_{Q\eta}(\eta - \hat{\eta}) + \phi(B_{QE}\gamma + B_{QQ}))(Q - \hat{Q})S$$
$$- \phi^2\gamma(B_{QE}\gamma + B_{QQ}))v(\eta)d\eta = 0 \tag{59}$$

Then apply the derivative $dQ/d\eta$ (equation (18)) to eliminate the $Q - \hat{Q}$ term:

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

Applying FOC_{μ_0} (equation (55)) again yields:

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

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 η conditional on the standard binding). We can then simplify the $B_{Q\eta}^2(\eta - \hat{\eta})^2$ term inside the integral in equation (61) as follows:

$$B_{Q\eta}^2 \int_{\hat{\eta}}^{\eta_H} (\eta - \hat{\eta})^2 v(\eta)d\eta = B_{Q\eta}^2 \int_{\hat{\eta}}^{\eta_H} (\eta - \dot{\eta} + \dot{\eta} - \hat{\eta})^2 v(\eta)d\eta$$
$$= (1 - V(\hat{\eta}))B_{Q\eta}^2(\sigma_{\eta c}^2 + (\dot{\eta} - \hat{\eta})^2)$$
$$= (1 - V(\hat{\eta})) \left(B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\frac{(B_{QE}\gamma + B_{QQ})^2}{(B_{QE} + B_{EE}\gamma)^2} \right) \tag{62}$$

where the last line makes use of equation (55) and the definition of $\dot{\eta}$, and $\sigma_{\eta c}^2$ denotes the variance of η conditional on the standard binding.

Finally, substitute equation (62) into equation (61) and combine terms to obtain:

$$\text{FOC}_\gamma : \frac{1 - V(\hat{\eta})}{S^2}[(B_{QE} + B_{EE}\gamma) \left(B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\frac{(B_{QE}\gamma + B_{QQ})^2}{(B_{QE} + B_{EE}\gamma)^2} \right) - \frac{\phi^2(B_{QE}\gamma + B_{QQ})^2}{B_{QE} + B_{EE}\gamma}$$
$$+ \phi^2\gamma(B_{QE}\gamma + B_{QQ})] = 0 \tag{63}$$

$$\Leftrightarrow \frac{1 - V(\hat{\eta})}{S^2}[(B_{QE} + B_{EE}\gamma)B_{Q\eta}^2\sigma_{\eta c}^2 + \phi^2\gamma(B_{QE}\gamma + B_{QQ})] = 0 \tag{64}$$

$$\Leftrightarrow \frac{(1 - V(\hat{\eta}))(B_{QE}\gamma + B_{QQ})\phi^2}{(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2} \left[\frac{B_{QE} + B_{EE}\gamma}{B_{QE}\gamma + B_{QQ}} \cdot \frac{B_{Q\eta}^2\sigma_{\eta c}^2}{\phi^2} + \gamma \right] = 0. \tag{65}$$

The solution to equation (65) requires a unique $\gamma \in (0, -B_{QE}/B_{EE})$.[43] The optimal standard is therefore output-based, though the optimal slope γ^* 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 γ inside the brackets in equation (65).

Equation (65) also produces the three comparative static results listed at the end of section 3.2. First, increasing the ratio $\sigma_{\eta c}^2/\phi^2$ increases the magnitude of the left-hand (and negative) term in brackets, requiring a larger value of γ to set equation (65) equal to zero. Second, increasing the ratio $-B_{QE}/B_{EE}$ increases the numerator relative to the magnitude of the denominator of the left-hand term in brackets, again requiring a larger value of γ

---

[43]For $\gamma \in [0, -B_{QE}/B_{EE}]$, the term in FOC (65) 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_γ is strictly positive at $\gamma = 0$ and strictly negative at $\gamma = -B_{QE}/B_{EE}$, implying an interior optimum. Furthermore, the optimal γ is unique, since the derivative of the term in brackets with respect to γ is strictly positive for $\gamma \in (0, -B_{QE}/B_{EE})$.

to set equation (65) equal to zero. Finally, to see the third comparative static regarding the inelasticity of demand for Q, 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 (18). 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 (65) by $(B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ})^2$ to see that, holding the variance of output fixed, the welfare-maximizing γ increases with the magnitude of $B_{EE}\gamma^2 + 2B_{QE}\gamma + B_{QQ}$.

A.5 Welfare-maximizing standard under uncertainty in η and F

Under uncertainty in both η and F, we can write the regulator's FOCs for μ_0 and γ as:

$$\text{FOC}_{\mu_0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \left(B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) w(F|\eta)v(\eta)dFd\eta = 0 \tag{66}$$

$$\text{FOC}_{\gamma} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} \left(B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) w(F|\eta)v(\eta)dFd\eta = 0, \tag{67}$$

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 (μ_0, γ) and a demand shock η.

Because deriving an analytic expression for γ^* is algebraically intractable, I instead evaluate the FOCs (66) and (67) at $\gamma = 0$, where the sign of FOC_{γ} informs the sign of γ^* (and is equivalent to the sign of γ^* assuming that the uniqueness property discussed in footnote 43 holds when F is stochastic).

Using expressions (22), (23), (24), and (25), FOCs (66) and (67) reduce to the following at $\gamma = 0$:

$$\text{FOC}_{\mu_0}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E - \phi)w(F|\eta)v(\eta)dFd\eta = 0 \tag{68}$$

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{\eta_L}^{\eta_H} \int_{F_L}^{\hat{F}} (B_E Q - \phi Q)w(F|\eta)v(\eta)dFd\eta = 0. \tag{69}$$

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 (68) 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 (69) as:

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

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:

$$B_E(Q, E, \eta, F) = \frac{-B_{QE}B_{Q\eta}(\eta - \bar{\eta})}{B_{QQ}} + B_{EF}(F - \hat{F}). \tag{71}$$

Substituting into equation (70) yields:

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

(72)

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

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

(73)

which implies:

$$\text{sign FOC}_\gamma|_{\gamma=0} = \text{sign} \left(E_c[(\eta - E_c[\eta])^2] - \frac{B_{QQ}B_{EF}}{B_{QE}B_{Q\eta}}E_c[(F - \hat{F}(E_c[\eta]))(\eta - E_c[\eta])] \right) \quad (74)$$

Equation (74) therefore demonstrates that the sign of $\text{FOC}_\gamma|_{\gamma=0}$, which in turn dictates the sign of γ^*, depends on the difference between the variance of η (conditional on the standard binding) and the covariance of η with F (again conditional on the standard binding, and weighted by $B_{QQ}B_{EF}/B_{QE}B_{Q\eta}$). Thus, if the correlation between η and F is sufficiently large, the optimal γ can be strictly negative. Also note that if F is non-stochastic, equation (73) reduces to equation (65), evaluated at $\gamma = 0$.

A.6 Attribute-basing a standard that is already output-based

This appendix assesses how the main result from section 3.1 and appendix A.3—that the optimal amount of attribute-basing is zero when there is only uncertainty in F—is affected when the standard in question is already output-based. Put another way, is there a case for attribute-basing a standard that is already output-based?

The economic environment now includes emissions E, output Q, and the attribute A. The planner's problem is to choose the μ_0 and γ parameters for an output and attribute-based standard given by $E \leq \mu_0 + \lambda Q + \gamma A$, where $\lambda > 0$ is pre-determined and the objective function is:

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

(75)

The agent's problem, when the standard binds, is to choose Q and A to maximize $B(Q, A, \mu_0 + \lambda Q + \gamma A, F)$. The agent's FOCs are:

$$B_Q(Q, A, \mu_0 + \lambda Q + \gamma A, F) = -\lambda B_E(Q, A, \mu_0 + \lambda Q + \gamma A, F) \quad (76)$$
$$B_A(Q, A, \mu_0 + \lambda Q + \gamma A, F) = -\gamma B_E(Q, A, \mu_0 + \lambda Q + \gamma A, F) \quad (77)$$

The implicit function theorem then leads to the following comparative statics when eval-

uated at $\gamma = 0$:[44]

$$\frac{dQ}{d\mu_0} = \frac{-1}{S} \left(\lambda B_{EE} B_{AA} + B_{QE} B_{AA} - \lambda B_{AE}^2 - B_{QA} B_{AE} \right) \tag{78}$$

$$\frac{dQ}{d\gamma} = \frac{dQ}{d\mu_0} A + \frac{B_E(Q, A, \mu_0 + \lambda Q, F)}{S} \left(\lambda B_{AE} + B_{QA} \right), \tag{79}$$

where S is the determinant of the Hessian and is assumed to be strictly greater than zero (so that the SOC holds).

The planner's FOC for μ_0, evaluated at $\gamma = 0$, is given by:

$$\text{FOC}_{\mu_0}|_{\gamma=0} : \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left(B_Q \frac{dQ}{d\mu_0} + B_E \frac{dE}{d\mu_0} - \phi \frac{dE}{d\mu_0} \right) w(F) dF = 0, \tag{80}$$

where B_Q and B_E are shorthand for $B_Q(Q, A, \mu_0 + \lambda Q, F)$ and $B_E(Q, A, \mu_0 + \lambda Q, F)$, and there is no term involving B_A because $B_A = 0$ at $\gamma = 0$, per equation (77). Now apply equation (76) and the fact that $dE/d\mu_0 = 1 + \lambda dQ/d\mu_0$ to obtain:

$$\text{FOC}_{\mu_0}|_{\gamma=0} : \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left(B_E(1 + \lambda \frac{dQ}{d\mu_0}) - \lambda B_E \frac{dQ}{d\mu_0} - \phi(1 + \lambda \frac{dQ}{d\mu_0}) \right) w(F) dF = 0 \tag{81}$$

$$\Leftrightarrow \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left(B_E - \phi - \phi \lambda \frac{dQ}{d\mu_0} \right) w(F) dF = 0. \tag{82}$$

The planner's FOC for γ, evaluated at $\gamma = 0$, is given by:

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left(B_Q \frac{dQ}{d\gamma} + B_E \frac{dE}{d\gamma} - \phi \frac{dE}{d\gamma} \right) w(F) dF = 0, \tag{83}$$

where again there is no term involving B_A because $B_A = 0$ at $\gamma = 0$. The sign of equation (83) dictates whether, and in what direction, the output-based standard should also be attribute-based.

Now apply equation (76) and the fact that $dE/d\gamma = A + \lambda dQ/d\gamma$ to obtain:

$$\text{FOC}_{\gamma}|_{\gamma=0} : \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left(B_E(A + \lambda \frac{dQ}{d\gamma}) - \lambda B_E \frac{dQ}{d\gamma} - \phi(A + \lambda \frac{dQ}{d\gamma}) \right) w(F) dF = 0 \tag{84}$$

$$\Leftrightarrow \int_{F_L}^{\hat{F}(\mu_0, \lambda, \gamma)} \left((B_E - \phi)A - \phi \lambda \frac{dQ}{d\gamma} \right) w(F) dF = 0. \tag{85}$$

---

[44]I focus on deriving results that hold at $\gamma = 0$ because the agent's comparative statics and then the planner's problem become intractable otherwise.

Finally, substituting equation (79) yields:

$$\text{FOC}_\gamma|_{\gamma=0} : \int_{F_L}^{\hat{F}(\mu_0,\lambda,\gamma)} \left(\left(B_E - \phi - \phi\lambda\frac{dQ}{d\mu_0} \right) A - \frac{\phi B_E}{S}(\lambda^2 B_{AE} + \lambda B_{QA}) \right) w(F)dF = 0.$$

(86)

A is constant at $\gamma = 0$, so the left term inside the integral in (86) therefore equals zero, assuming μ_0 is chosen optimally per equation (82). $\phi B_E/S$ is strictly positive, so the sign of the optimal gamma is given by the sign of $-(\lambda^2 B_{AE} + \lambda B_{QA})$.

First, note that if $\lambda = 0$, so that the standard is not output-based, the optimal γ then equals zero as well, agreeing with the original results in section 3.1.

Now, suppose $\lambda > 0$ and that either $B_{AE} > 0$ or $B_{QA} > 0$. In the fuel economy context, the former holds if emissions abatement is costlier for larger vehicles, and the latter holds if the marginal utility of miles traveled is greater for larger vehicles. In either of these cases, equation (86) tells us that the optimal γ is strictly less than zero. That is, the optimal standard is reverse attribute-based.

The intuition for this result follows the theory of the second-best. Output-basing ($\lambda > 0$) creates a distortion to Q, particularly for low realizations of F. If $B_{AE} > 0$, then reverse attribute-basing with $\gamma < 0$ will cause agents to reduce E, and therefore Q, for low realizations of F, thereby partially mitigating the distortion to Q. If $B_{QA} > 0$, there is an even more direct benefit to reverse attribute-basing, since forcing agents to choose a lower A at low realizations of F directly reduces their marginal value of Q.

B   Additional detail for numerical calibration

B.1   Footprint-based fuel economy standards

To calibrate B_{QQ}, I begin with table 3, column (2) in Ito and Sallee (2018), which provides
an estimate of B_{QQ} of -\$0.230/kg^2. To convert this estimate from \$ per kg^2 to \$ per ft^4,
I use Whitefoot and Skerlos (2012), which 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 56.8 ft^2 and weighed 3666 lb. Multiplying Ito and Sallee's (2018)
-\$0.230/kg^2 by the square of this weight-to-footprint ratio yields the B_{QQ} of -\$197/ft^4 in
table 2.

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. I follow the procedure discussed in Kellogg (2018) to derive B_{EE} = -\$1756
per (gallon/100 miles)^2 from National Research Council (2015).

I derive B_{QE} from the fact that the actual slope parameter γ for U.S. fuel economy
regulations was drawn to match $-B_{QE}/B_{EE}$, the rate at which private agents increase E
with Q. 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 per (gal·ft^2/100mi) in table 2.

When calculating the value of B_{EF}, I use the new vehicle data from Busse, Knittel, and
Zettelmeyer (2013) that were originally sourced from the National Highway Transportation
Survey.

For the model's initial condition in 2012, the average U.S. footprint and fuel economy (Q_0
and E_0) were 56.8 ft^2 and 4.01 gal/100mi, respectively (Leard, Linn, and McConnell 2017).
I then compute $\bar{Q}$ and $\bar{E}$—the baseline 2012 footprint and fuel economy that would have
been chosen in the absence of regulation—by combining the model with data from Leard and
McConnell (2017) on the shadow value of the 2012 fuel economy constraint.[45] To compute $\bar{Q}$
and $\bar{E}$ from Q_0 and E_0, first let ϕ_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 $\bar{Q} = Q_0 + (\bar{E} - E_0)D$. The resulting $\bar{Q}$ and $\bar{E}$ are 56.8 ft^2 and 4.37 gal/100mi, respectively.

To compute F_0 and σ_F, I use the tax-inclusive all grades and all formulations retail gaso-
line price series from the EIA,[46] deflated to \$2012 by the Bureau of Labor Statistic's CPI for
all urban consumers, all items less energy, not seasonally adjusted (series CUUR0000SA0LE).
The three-year moving average price that I calculate for each month t uses data from t and
from the prior 35 months. I calculate t-month differences in this moving average using
observations from January, 2004 (the first month for which a ten-year difference can be cal-
culated) through 2012. Volatility increases with the time horizon: the one-month volatility
is \$0.02/gal, while the ten-year volatility is \$0.46/gal. I set σ_F equal to the average volatility

---

[45]Leard and McConnell (2017) derive a shadow price of \$36 per metric ton of CO_2 using information on
the price of tradeable fuel economy credits from Tesla Motors' 2013 SEC 10-K filing.

[46]The EIA gasoline price data are available at https://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_nus_a.htm.

across all horizons $t \in [1, 120]$, yielding $\sigma_F = \$0.35/\text{gal}$.

To estimate σ_Q, I first estimate an AR1 process for Q using the annual footprint data from Leard, Linn, and McConnell (2017). I use only data through 2010, since after that time fuel economy standards for cars became attribute-based and started increasing in stringency. The process I estimate includes a time trend and results in an estimated AR1 coefficient $\hat{\rho} = 0.48$ and a root mean squared error $\sigma_\varepsilon = 0.16 \text{ ft}^2$.[47] Because the time series of footprints only includes 15 observations, I use the bias correction originally proposed by Orcutt and Winokur (1969) and suggested by MacKinnon and Smith (1998), and therefore set $\rho = (T\hat{\rho} + 1)/(T - 3) = 0.68$. The volatility of footprint at a horizon of t years is then given by $\sigma_\varepsilon \sqrt{(1 - \rho^{2t})/(1 - \rho^2)}$. Calculating this expression for each year $t = 1, 2, ...10$ and then averaging over ten years yields $\sigma_Q = 0.21 \text{ ft}^2$.

In an alternative specification, I assume a random walk with drift for Q. Under that assumption, the standard deviation of first differences in Q is 0.17 ft^2, and averaging volatility over a projected ten-year horizon yields $\sigma_Q = 0.39 \text{ ft}^2$.

B.2 Fuel economy standards with endogenous miles traveled

I estimate σ_Q using historical data on average U.S. vehicle miles traveled (VMT) from the Federal Reserve Bank of St. Louis.[48] 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 26) and dividing by total miles traveled in 2012.

Because shocks to fuel prices F induce changes to miles traveled, I must adjust the miles traveled time series for these shocks in order to isolate volatility induced by η. 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 $\$3.69/\text{gallon}$.[49] Let $\hat{Q}$ denote the time series of adjusted miles traveled.

I project annual volatility in $\hat{Q}$ out over a 10-year policy horizon under the assumption that $\hat{Q}$ follows a random walk with drift.[50] I first compute annual volatility by taking the standard deviation of first-differences in the $\hat{Q}$ series. The volatility of $\hat{Q}$ at a horizon of t years is then given by the product of the annual volatility with $\sqrt{t}$.[51] I compute volatility for each year $t = 1, 2, ...10$ and then average over ten years, yielding $\sigma_Q = 3517$ miles.

---

[47] When I test for a unit root using the augmented Dickey-Fuller test of Elliot, Rothenberg, and Stock (1996), I obtain a test statistic of -3.08 at the Ng and Perron (2001) optimal lag length of 4, relative to 10% and 5% critical values of -2.29 and -3.01, respectively. I search over lag lengths starting from a maximum of 4 lags, since the 7 lags implied by the Schwert (1989) criterion leave too few remaining observations to run the estimator.

[48] Annual U.S. VMT data were accessed from https://fred.stlouisfed.org/series/TRFVOLUSM227NFWA on 2 May, 2018

[49] 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.1, equation (19), where γ_0 denotes the 2012 average fleet-wide fuel economy of 4.01 gallons per 100 miles.

[50] The AR1 coefficient from a regression of miles traveled on lagged miles traveled and a time trend is 1.04.

[51] Note that t is the limit, as ρ goes to 1, of the $(1 - \rho^{2t})/(1 - \rho^2)$ expression for the AR1 process from appendix B.1.

B.3 Electricity sector emissions standards

To calculate B_{QQ}, I use the fact that $dQ/dP = 1/B_{QQ}$. To translate the electricity demand elasticity to dQ/dP, I use the value of U.S. total power consumption for 2015 of $\bar{Q} = 4078$ TWh and the 2015 average retail price to residential end users of \$0.127/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. That table is available at https://www.eia.gov/electricity/annual/archive/03482016.pdf.

To calibrate B_{EE}, I use Cullen and Mansur's (2017) estimate that a \$40 per ton tax on CO_2 will reduce emissions by 7.9%, holding electricity output fixed, and the relationship $dE/d\phi = 1/B_{EE}$. To translate Cullen and Mansur's (2017) percentage result to a result in levels, I use $\bar{E} = 2031$ 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.

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. 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 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 19.15 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 below).

The monthly front-month natural gas futures prices that I use to calculate σ_F are available from the EIA 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 below) because those prices are available at a monthly frequency only back to 2002.

To estimate σ_Q, I use national-level electric generation data that are available from the EIA at https://www.eia.gov/electricity/data/state/annual_generation_state.xls. I must first adjust the generation time series for changes induced by fluctuations in the price of natural gas. I use annual EIA data on the purchase price of natural gas for the electric power sector, covering 1997–2018, and available at https://www.eia.gov/dnav/ng/hist/n3045us3A.htm. I deflate these prices 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 (13), for unconstrained agents, in appendix A.1.

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 (13), the direct effect of F on Q in the data is incorporated into my estimate of σ_η. Though this direct effect should induce a positive

correlation between F $(= -P_g)$ and η, I estimate an overall small negative correlation of -0.06, since shocks to power demand are on average negatively correlated with F. In the calibration I enforce $\rho = 0$.

I project annual volatility in adjusted generation $\hat{Q}$ out over a 15-year policy horizon under the assumption that $\hat{Q}$ follows a random walk with drift.[52] My computation of this projection follows the same procedure that I used in the miles traveled application, discussed in appendix B.2.

References

[1] Elliot, Graham, Thomas J. Rothenberg, and James H. Stock (1996), "Efficient Tests for an Autoregressive Unit Root", *Econometrica* 64(4), 813-836.

[2] 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.

[3] MacKinnon, James G. and Anthony A. Smith, Jr. (1998), "Approximate Bias Correction in Econometrics", *Journal of Econometrics* 85, 205-230.

[4] Ng, Serena and Pierre Perron (2001), "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power", *Econometrica* 69(6), 1519–1554.

[5] Orcutt, Guy H. and Herbert S. Winokur (1969), "First Order Autoregression: Inference, Estimation, and Prediction", *Econometrica* 37(1), 1-14.

[6] Schwert, G. William (1989), "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business & Economic Statistics* 7(2), 5-17.

[7] 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.

---

[52]The AR1 coefficient from a regression of generation on lagged generation and a time trend is 0.62. After applying the correction $\rho = (T\hat{\rho} + 1)/(T - 3) =$ to account for the short (19 period) generation time series, the AR1 coefficient is 0.80. When I test for a unit root using the augmented Dickey-Fuller test of Elliot, Rothenberg, and Stock (1996), I obtain a test statistic of -2.19 at the Ng and Perron (2001) optimal lag length of 0, relative to the 10% critical value of -3.20. I search over lag lengths starting from the maximum of 7 lags, per the Schwert (1989) criterion.