Pigou Creates Losers: On the Implausibility of Achieving Pareto Improvements from Efficiency-Enhancing Policies

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

September 12, 2022

Abstract

Economic theory predicts that efficiency-enhancing policy changes can be made to benefit everyone through the use of lump-sum transfers that compensate anyone initially harmed by the change. Precise targeting of compensating transfers, however, may not be possible when agents are heterogeneous and the planner faces constraints on the design of transfers, due, for example, to asymmetric information. In this paper, I derive an impossibility condition showing when Pareto improvements are not possible. The condition can be directly tested with readily available data. It relates the size of efficiency gains to the degree of predictability between initial burdens and variables used to condition transfers. The main empirical application is to a gasoline tax to correct carbon emissions, but I present related results for other sin taxes. Results indicate that it is infeasible to create a Pareto improvement from the taxation of these goods, and moreover that plausible policies are likely to leave a large fraction of households as net losers. The paper argues that the existence of these losers is relevant to policy design and may help explain political challenges faced by many efficient policies.

Keywords: Corrective taxation, externalities, equity
JEL: H23, Q58, L51

*The author thanks Karl Dunkle-Warner, Catherine Wright, Marshall Blundell and Sophie Andrews for excellent research assistance, Shanjun Li and Gilbert Metcalf for conference discussion, and Lint Barrage, Severin Borenstein, Lucas Davis, Don Fullerton, Kelsey Jack, Gary Libecap, Ethan Ligon, Dmitry Taubinsky and seminar participants at Berkeley, Brown, the CBO, Claremont McKenna, Davis, HKUST, as well as conference participants at ASSA, NBER, NTA and HERE for helpful comments. The Sloan Foundation and the Hellman Fund at the University of California, Berkeley provided generous support. Contact: sallee@berkeley.edu.

1 Introduction

Why do efficient policies so often fail to gain political traction? Many policies are widely viewed as desirable by economists but unpopular with the public and unsuccessful in the policy process. Examples range from the pricing of pollution to repeal of the mortgage interest tax deduction to free trade.

Several factors may lead to unpopularity of such policies, one of which regards the distribution of burdens they induce. Distributional concerns come in two varieties. In one, a policy is disliked because it is regressive and disproportionately affects low-income households. In the other, a policy imposes a substantial burden on a particular set of stakeholders who mobilize to block the policy.

In either case, economic theory provides a potential reply, which is that any such losers can be compensated. Any efficiency-enhancing policy, by definition, creates enough new surplus to compensate all losers. That is, any Kaldor-Hicks efficiency gain can be made into a Pareto improvement, if the right transfers are made in the background. A regressive tax could be combined with tax reform so as to preserve the desired income distribution, or any firms facing lost profits can be made whole.

Economists are well aware that, in practice, these transfers are often not made for various reasons. The point of this paper is to make clear that it is often not even *possible* to design transfers that achieve a Pareto improvement, even if the political will and administrative capacity exists. The reason is that transfer schemes are inherently constrained. They must be based on a set of observable characteristics, which will only be imperfectly correlated with burdens. One reason for this is asymmetric information, but constraints may also arise from demands for parsimony, administrative feasibility, or other factors.

In this paper, I develop a method for determining whether Pareto improvements are impossible that can be applied to a range of contexts. I derive a theoretical model that yields an impossibility result that indicates conditions under which a Pareto improvement is not possible because of imperfections in the targeting of transfers. This condition can be taken directly to data. Empirically, I consider the case of externality-correcting taxes in the US, with a focus on motor fuels, when transfers are based on household demographics, income and geography. I find that a Pareto improvement is impossible, and that a substantial fraction of households will inevitably be net losers from any externality-correcting policies, even accounting for revenue recycling. In brief, Pigouvian taxes create losers.

The basic idea is best illustrated via example. Consider a tax that increases efficiency by correcting an unpriced externality in the tradition of Pigou (1932). This policy creates a heterogeneous initial distribution of burdens across individuals depending on their taste for the taxed good. The planner has enough revenue collected from the tax in order to compensate everyone for their loss through lump-sum transfers.[1] But, compensating everyone will require giving back the transfers in a targeted way. Targeting directly on consumption of the good itself will undo the desired corrective incentives, so the transfer must be based on factors like demographics, geography or income. If the transfer function is not rich enough to precisely target transfers, then the planner will run out of available funds before fully compensating everyone. In this sense, the failure to create a Pareto improvement is due to a *prediction* problem; lump-sum transfers can only undo the distribution of burdens if they can be targeted precisely.

Summary of the paper: The paper first describes a theoretical framework for analyzing efficiency-enhancing policy changes and then derives an impossibility condition. I show that a Pareto improvement is impossible unless the variables (which I call covariates) that are used to determine the transfer scheme can predict initial policy burdens with sufficient precision. The degree of precision required is a simple function of the size of the average surplus gain created from the efficiency-enhancing action. I first derive an impossibility condition when the covariates are assumed to be exogenous. I then show that the condition holds with only a slight modification when the covariates are endogenous so that the transfer scheme can create distortions.

The impossibility condition can be directly tested with data, with the exact data requirements dependent on the policy in question. For a marginal increase in an externality-correcting tax, the initial burdens are measured directly by baseline consumption of the good, and the average welfare gain depends on an estimate of marginal external damages and a demand derivative. Thus, to check the condition for an externality-correcting tax, one needs (1) an estimate of the distribution of baseline consumption of the good, (2) knowledge of the correlation between baseline consumption and covariates that can be used in a transfer scheme, (3) an estimate of the own-price derivative, and (4) an estimate of the size of the externality.

To take the theory to data, I use the Consumer Expenditure Survey (CEX) to estimate the distribution of consumption of externality-creating goods and the correlation between consumption and covariates that could be used in transfer schemes. I combine this with estimates from the literature of the size of externalities and price derivatives. I initially focus on a gasoline tax used to correct carbon-related externalities. There is wide dispersion in consumption of gasoline across households, and only a modest fraction of this variation

[1] To be precise, compensation may require taxing those who gained from the externality reduction, not just recycling of the revenue. For illustration in this example, I assume the case where revenue is sufficient to cover the losses, but the actual model is more general.

is correlated with variables that are likely to influence a transfer scheme, namely household structure, geographic location and income. Only about one-third of intrahousehold variation in annual gasoline expenditures is predictable by those variables, based on OLS and lasso models. Using conventional estimates of the externality gain achieved by a carbon tax, I conclude that the transfer scheme is nowhere close to precise enough to create a Pareto improvement.

I then show that the degree of predictability is no better for other externality-causing goods measured in the CEX, namely natural gas, electricity, alcohol and tobacco. I address measurement error concerns with the use of two auxiliary data sets. I interpret this as evidence that it will be infeasible to create a Pareto improvement from corrective taxes on these goods, even when a planner uses an implausible amount of information to create an unrealistically flexible lump-sum transfer scheme. One potential exception to this conclusion is if past consumption can be used as a predictor of future consumption. I explore this possibility with additional panel data on electricity consumption in California.

Contributions and relationship to the literature: A tradition in economics going back at least to Musgrave (1959) suggests that efficiency and equity concerns can often be conceptually divided. Given tools that can tilt the balance between rich and poor, like a progressive income tax, a policymaker should ensure market efficiency, and then simply dial up (or down) the levers that determine the income distribution to achieve the desired resource allocation in society. This is an extremely useful modeling device, and it is favored by many who study second-best tax design (e.g., Kaplow 2004). A literature in public finance explores the separability of efficiency-enhancing policies, including Pigouvian taxes, in second-best constrained environments (e.g., Gauthier and Laroque 2009; Kaplow 2012). This theoretical literature has noted that preference homogeneity is a critical assumption in their models, but little empirical work follows up by asking how these ideas can be implemented when there is some heterogeneity. Closely related is a seminal result of optimal tax theory that the distributional implications of a commodity tax are irrelevant in the presence of a nonlinear income tax (Atkinson and Stiglitz 1976). This likewise requires preference homogeneity (Saez 2002). This paper comments on these theoretical traditions by (1) deriving theoretical conditions that demonstrate when Pareto improvements are impossible in the presence of some preference heterogeneity, (2) empirically testing the degree to which transfers can be adequately targeted so as to undo the initial distributional burdens of a class of policies, and (3) demonstrating the relationship between heterogeneity and empirical prediction in achieving separation.

This paper bears an apparent relationship to several strains of literature in the theory of taxation, but it ultimately deals with different concerns. First, in being concerned with the correlation of tax burdens with covariates, the paper is related to the literature on tagging and targeting that follows Akerlof (1978), which considers how observable characteristics can be used to reduce distortionary tax incentives. Second, in being driven by a root information problem, this paper bears some relation to the literature begun by Mirrlees (1971), in which, if the planner could directly observe everyone's ability level, the optimal tax system would be nondistortionary. In my setting, if the planner could directly observe preference heterogeneity (and all other primitives that determine consumption of the externality-causing good), then Pareto improvements will be straightforward. Third, in considering optimal tax and transfer schemes to correct externalities, this paper is related to a literature—starting with the seminal work of Sandmo (1975) and with key contributions including Bovenberg and van der Ploeg (1994); Cremer, Gahvari, and Ladoux (1998, 2003); Jacobs and de Mooij (2015)—that derives second-best taxes on externality-creating commodities in order to maximize social welfare.

All three of these literatures are focused on how to derive second-best policies that minimize distortions caused by tax and transfer systems. My objective is different, at least proximately. My goal is to characterize the ways that imperfect information, which results in imperfect targeting/tagging, limits the planner's control over the final distribution of outcomes induced by an efficiency-enhancing reform. My empirical exercise is closer to the literature on targeting on observables prominent in the development literature, where the goal is to use readily measured proxies for wealth to target social programs. (See Coady, Grosh, and Hoddinott (2004) for a review.) The question at hand in designing the lump-sum transfer schemes is not maximization of social welfare (though that is the deeper reason why efficiency-enhancing policies are undertaken to begin with), but rather how to compensate the losers from the efficient scheme, with an eye on political economy, as explained next.

Should we be concerned about creating a Pareto improvement, or is it a red herring? Pareto efficiency vis-à-vis the status quo is quite distinct from social welfare maximization. If one begins with the objective of maximizing social welfare, there is no reason to prioritize the status quo resource allocation in society, so fussing over Pareto improvements is largely a distraction. The motivation for seeking Pareto improvements in this paper is instead a practical one. The political process tends to favor the status quo over changes, and as such, effecting change requires satisfying a great many people. That is, a utilitarian planner would gladly accept a policy that benefits most people, but causes modest harm to the remainder. But, in practical terms, even small numbers of losers can create substantial political obstacles, consistent with the logic of collective action (Olson 1965, 1982). Empirically, this paper suggests that even implausibly well-designed schemes will leave large fractions of households as net losers. Economists should not assume that it is straightforward to compensate all losers from a reform, even where political will and administrative capacity exists. They should instead view net losers as an inevitable by-product of efficiency-enhancing reform that require consideration in the policy-making process.

In terms of the empirical application, this paper contributes to an existing literature on the distributional impacts of gasoline taxes (e.g., Poterba 1991; West 2004) and carbon taxes (e.g., Hassett, Mathur, and Metcalf 2009; Grainger and Kolstad 2010; Dinan 2012; Mathur and Morris 2014; Metcalf 2009; Burtraw, Sweeney, and Walls 2008; Williams, Gordon, Burtraw, Carbone, and Morgenstern 2015). That work has been overwhelmingly focused on measuring *average regressivity* (or progressivity) of taxes, whereas this paper is sharply focused on heterogeneity in policy burdens conditional on income and the degree to which that heterogeneity can be controlled via a transfer scheme.

A smaller recent literature does quantify heterogeneity in policy burdens conditional on income. Rausch, Metcalf, and Reilly (2011) use the Consumer Expenditure Survey (CEX) to characterize the overall progressivity of carbon pricing, accounting for both consumption and income channels. Pizer and Sexton (2019) analyze the CEX and similar data from the United Kingdom and Mexico to show box plots that depict the range of energy consumption within income deciles. Fischer and Pizer (2019) explore how attention to horizontal equity influences a comparison between energy-pricing schemes and a performance standard. Cronin, Fullerton, and Sexton (2019) link the CEX to income tax data to explore a variety of revenue recycling mechanisms and quantify the variation in burdens that remains, taking into account fine-grained differences in income sources. Davis and Knittel (2019) show the heterogeneity in policy impacts of fuel-economy standards across different households in the same income decile in what is otherwise a study of average progressivity.[2]

These papers provide several initial results that are important for the development of a full analysis of heterogeneity in the incidence of energy policies. All demonstrate that there is significant heterogeneity in baseline energy consumption within households that have similar income, which are consistent with the descriptive facts I document here. Only Cronin, Fullerton, and Sexton (2019) link their study of heterogeneity to revenue redistribution schemes. They model several realistic schemes for revenue redistribution using detailed administrative tax records to show how the distribution of burdens depends on the use of revenue. I complement their approach first by modeling alternative transfer schemes that are explicitly designed to reduce heterogeneity in burdens, and second by providing a theoretical framework that demonstrates under what conditions revenue redistribution cannot support

[2] Some of this literature invokes the concept of horizontal equity. This concept has come under criticism as a normative criterion (Kaplow 1989). Here, I am concerned with horizontal equity, but not for normative reasons. This paper's argument is that horizontal equity matters for preventing the creation of losers and thereby improving political acceptability.

a Pareto improvement.

Many prior studies have discussed compensation schemes from externality-correcting taxes, and careful writers do sometimes note that schemes that achieve average redistributional goals will nevertheless create some losers (e.g., Metcalf 2018, p.98). Cronin, Fullerton, and Sexton (2019) and Fischer and Pizer (2019) both conjecture that, when there is a great deal of heterogeneity in baseline energy usage, it will be impossible to design transfer schemes to make everyone better off. My model offers a way to confirm those conjectures, and to show how much heterogeneity is requied to rule out true Pareto transfers.[3] I turn now to a description of that model.

2 Theory: A model of Pareto transfers

Costs, benefits and revenue: Consider some policy action that will create heterogeneous costs, produce efficiency gains, and raise some revenue that will be redistributed back to agents. Heterogeneous agents are indexed $1 = 1, ..., N$.

Initial costs for each agent are denoted $\tilde{c}_i$ measured in dollars (not utils), with $\sum_i \tilde{c}_i = C > 0$ being the aggregate initial cost of the policy. Initial costs c_i can be positive, negative or zero for an individual, but aggregate cost is assumed to be positive. These costs are initial in the sense of being considered before transfers are allocated, but they take into account behavioral responses the agent takes in response to policy. Denote the cost of the policy before behavioral responses as c_i, and define the net welfare gain from behavioral response as $b_i = c_i - \tilde{c}_i$. The average of b_i is denoted $\bar{b} = (\sum_i b_i)/N$.

For example, if the policy action were a pollution tax, c_i is the increase in expenditure on the good given the original level of consumption, and $\tilde{c}_i$ is the compensating variation for agent i associated with the price change resulting from the tax after taking their consumption response into account. By the envelope condition, these would be approximately equal ($b_i = 0$) for a small tax.

The policy action yields efficiency gains g_i, with $\sum_i g_i = G > 0$, measured in dollars. For the example of a pollution tax, g_i are the direct health or other welfare benefits of pollution reduction. Gains are assumed to be weakly positive ($g_i \geq 0 \ \forall i$). This assumption is not intended to be economically substantive, but it is used in the algebraic proofs below and

[3] Another related strand of literature focuses on compensating producers who are harmed by environmental regulation (Bovenberg and Goulder 2001; Bovenberg, Goulder, and Gurney 2005; Goulder, Hafstead, and Dworsky 2010). Most of that literature focuses on average impacts by sector or consumer group and does not delve into the heterogeneity that is the core of this study, though Burtraw and Palmer (2008) do consider individual power plants in an examination of the impacts on the electricity sector.

thus is an important restriction. Denote the average efficiency gain as $\bar{g} = G/N$.[4]

The policy raises some revenue, denoted $R > 0$. Revenue can be redistributed through a transfer scheme based on a vector of covariates, $\mathbf{X}_i$. The transfer scheme is denoted $T(\mathbf{X}_i)$. The budget constraint requires that total transfers given out to agents is no greater than revenue $\sum_i T(\mathbf{X}_i) \leq R$. Note that $T(\mathbf{X}_i)$ can take on negative values—that is, the transfer can be a tax for some individuals. The average *funding gap* is the per person difference between revenue raised and cost, denoted $\bar{\Delta} \equiv (C - R)/N$. This gap can be positive, negative or zero. A positive gap implies that the policy imposes costs that exceed revenue, as would be the case for a pollution tax.

Here, I assume that $\mathbf{X}_i$ are exogenous characteristics, so the transfer function does not determine X_i. I call this case the "prediction interpretation." I relax this assumption in section 2.2, which I refer to as the "mechanism design interpretation."

There are two different interpretations about how the transfer function determines behavior that are both consistent with the algebra of the proof. One interpretation is that the policy and transfer function $T(\cdot)$ are implemented together and we can interpret G and $\tilde{c}_i$ as being the end result of behaviors resulting from the transfer. For a pollution tax, this means that G and $\tilde{c}_i$ take into account income effects of redistributing revenue.

Alternatively, one might assume that the transfer received by i does not determine G or c_i, and think of the effects of policy as being separable from the design of the transfer. This is more limiting, but may be more intuitive. For example, in the case of a pollution tax, this is equivalent to assuming first that income effects on pollution generation are uniform across agents, so that how the revenue is redistributed does not change the level of pollution, and second that the compensating variation associated with a price change does not depend locally on income within the range of variation created by the transfer.

A Pareto improvement: An efficiency-enhancing policy is defined as one in which total benefits plus revenue exceed total costs: $R + G > C$. This accords with the standard Kaldor-Hicks definition. A Pareto improvement occurs when the analogous condition holds at the individual level for each individual, not just on average. Thus, including the budget constraint, a Pareto improvement occurs when:

$$\tilde{c}_i - T(\mathbf{X}_i) \leq g_i \ \forall i \quad \text{and} \quad \sum_i T(\mathbf{X}_i) \leq R.$$

[4]Some benefits of a policy may go to agents that are outside the jurisdiction of the policymaker (e.g., future generations). Let $\eta\tilde{G} = G$ represent the full benefit, where η is the fraction of all benefits that accrue to agents $i = 1, ..., N$. The model implicitly assumes that all relevant agents that bear initial costs are included in the set $1, ..., N$, but that some benefits may accrue to others. This can be relaxed by assuming $\eta = 1$ and all those who benefit are included in the set $i = 1, ..., N$, allowing some of them to have $c_i = 0$.

To achieve a Pareto improvement, one must design a transfer scheme that delivers bigger transfers to those with bigger initial burdens. Intuitively, this requires transfers be targeted to offset burdens. Accordingly, I refer to $\tilde{c}_i - T(\mathbf{X_i})$ as the *targeting error*.

2.1 Impossibility condition (prediction interpretation)

The main result of the paper is a condition establishing when a Pareto improvement from an efficiency-enhancing policy action is impossible. Roughly, the condition states that, if targeting errors are large relative to efficiency gains, then it will be impossible to achieve a Pareto improvement.

Condition 1. *Let $\tilde{c}_i$ be the initial costs from a policy, N be the number of agents, $\mathbf{X}_i$ be a vector of exogenous covariates observed by the planner, $T(\mathbf{X}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, and $\bar{g}$ be the average efficiency gain. If the average absolute targeting error exceeds twice the average efficiency gain minus the average funding gap; i.e.,*

$$\frac{1}{N}\sum_i |\tilde{c}_i - T(\mathbf{X}_i)| > 2\bar{g} - \bar{\Delta},$$

then there is no distribution of g with $g_i \geq 0\ \forall i$ for which the policy and transfers create a Pareto improvement.

Condition 1 illustrates the relationship between the size of efficiency gains from the policy action and the ability of a policy to precisely target transfers based on initial costs. The left-hand size of the inequality is the average size of targeting errors. If these are "too large," a Pareto improvement will be impossible. How large is "too large" depends on the size of the efficiency gains, $\bar{g}$, as well as the size of the budget relative to the total amount of burdens created, which is summarized in $\bar{\Delta}$. As efficiency gains minus the budget gap grow smaller, the "margin of error" for targeting errors shrinks. If this margin of error gets too small, a Pareto improvement will be impossible. The proof of condition 1 is in the appendix.

Put another way, an efficiency-enhancing policy, by definition, creates some surplus. If transfers could be targeted perfectly, then a Pareto improvement would always be possible. With imperfect targeting, some agents will be overcompensated and others will be undercompensated. The condition determines how much misallocation can be tolerated and still have a *possibility* of creating a Pareto improvement. The setup is fully agnostic about the distribution of g. The derivation asks when there is enough surplus to compensate all losers, if by happy coincidence *gains are distributed perfectly so as to offset losses net of the transfers*. The impossibility condition shows when there is no distribution of the gains that could generate a Pareto improvement. The derivation with remains agnostic about the distribution of gains but assumes an ability to measure private costs is motivated by empirical application, as I explain below.

Condition 1 starts with data on individual costs net of behavioral responses, $\tilde{c}_i$. Data may be more readily available on c_i, the costs before behavioral responses are taken into account. Condition 2 restates the same condition but using c_i by introducing one additional term, $\bar{b}$, to the right-hand side of the inequality.

Condition 2. *Let c_i be the initial costs from a policy assuming no behavioral responses, N be the number of agents, $\mathbf{X}_i$ be a vector of exogenous covariates observed by the planner, $T(\mathbf{X}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, $\bar{g}$ be the average efficiency gain, and $\bar{b}$ be the average private welfare gain from behavioral responses. If the average absolute targeting error exceeds twice the sum of the average efficiency gain and behavioral adjustment gain minus the average funding gap; i.e.,*

$$\frac{1}{N}\sum_i |c_i - T(\mathbf{X}_i)| > 2(\bar{g} + \bar{b}) - \bar{\Delta},$$

then there is no distribution of g with $g_i \geq 0$ $\forall i$ for which the policy and transfers create a Pareto improvement.

Taking the condition to data: This condition was constructed with the aim of facilitating empirical application. To see how to apply the condition to data, consider the case of a gasoline tax targeting greenhouse gas emissions, which I develop further below. For a small change in the gas tax, the initial costs $\tilde{c}_i$ for agents are well approximated by the quantity of gasoline consumed (via Roy's identity), which is readily measured in survey data. The same survey data contains covariates, like income, that might be used to design transfers, so the data can be used to estimate targeting errors. The impossibility condition is written for a generic transfer function $T(\cdot)$. To take this to data, we simply search for a transfer function that minimizes the average absolute targeting error, understanding that this is the best case scenario. Thus, estimating the minimum size of the left-hand side of the inequality in condition 1 comes from predicting c_i with $\mathbf{X}_i$.

Given an elasticity of demand for gasoline and an estimate of the marginal externality, the average welfare gains and revenue gap can be estimated. Information about the *distribution* of efficiency gains (who exactly benefits from reduced emissions) is much harder to measure, but the condition is constructed so that this is not necessary. The condition shows when a Pareto improvement is impossible, regardless of how the efficiency gains are spread across agents.

Information about many other efficiency enhancing policy actions are likely to have the same feature, where some information about the joint distribution of burdens and covariates is available, along with a credible way of calculating the average efficiency gains, but not necessarily information about the joint distribution of those efficiency gains and initial costs. Given a different set of information, alternative conditions could be constructed.

2.2 Impossibility condition (mechanism design interpretation)

A key assumption of the model above is that the covariates in $\mathbf{X}$ are assumed to be exogenous. I call this the "prediction interpretation" because, as a result of that assumption, the main concern in the empirical application turns out to be straightforward prediction.

An alternative is to model the covariates $\mathbf{X}$ as endogenous. Then, the choice of the transfer function $T(\cdot)$ will create distortions as agents alter their choices of $\mathbf{X}$ in order to optimize the transfer they receive. I call this the "mechanism design interpretation" because the agent is assumed to have information about $\tilde{c}_i$ that is hidden to the planner, who sees only the covariates, which the agent can manipulate.

Specifically, define $\mathbf{X}_i$ as the initial vector of covariates of an agent that would result after the policy action was implemented, assuming an equal lump-sum transfer to all agents. Thus, $\mathbf{X}_i$ are the covariates (e.g., income) that the agent privately prefers if there is no incentive associated with a transfer function. The agent can, however, deviate to some other vector of covariates $\mathbf{X}'_i$ at a cost, where the cost is assumed to be rising in the difference between $\mathbf{X}_i$ and $\mathbf{X}'_i$. Specifically, I assume that there is some cost function (in dollars) measured by $\tau(|\mathbf{X}'_i - \mathbf{X}_i|)$, with $\tau(0) = 0$, $\tau'_j > 0, \tau''_{jj} > 0$ and $\tau''_{jk} \forall j \neq k$ where j and k index the variables in the vector. This is a convenient set of assumptions that says the cost of changing each covariate is symmetric, increasing and convex, and that changing one characteristic has no impact on the cost of changing others.

Then, for any given transfer function $T(\cdot)$, agent i will maximize welfare W_i:

$$\max_{\mathbf{X}'_i} W_i = -c_i + g_i + T(\mathbf{X}'_i) - \tau(|\mathbf{X}'_i - \mathbf{X}_i|). \tag{1}$$

Denote the vector of covariates that solves this problem for agent i as $\tilde{\mathbf{X}}_\mathbf{i}$. Note that $\tilde{\mathbf{X}}_\mathbf{i}$ will depend on the particular transfer function T. As in the "prediction interpretation" case, I maintain the assumption that G and c_i are independent of how revenue is recycled (or interpret G and c_i as resulting from a given $T(\cdot)$), but here add the same simplifying assumption that those variables are also independent of $\tilde{\mathbf{X}}$ (or they are interpreted as resulting from a given $T(\cdot)$). For the case of a pollution tax, this is equivalent to saying that distortions to income, for example, that result from the incentives created by the transfer function do not change the overall level of pollution or the compensating variation associated with the tax.

Denote the welfare cost τ that results for agent i from transfer scheme $T(\cdot)$ as τ^T, and let the mean of this be $\bar{\tau}^T = \sum_i \tau_i^T / N$.

Condition 3. *Let $\tilde{c}_i$ be the initial costs from a policy, N be the number of agents, $\tilde{\mathbf{X}}_i$ be a vector of endogenous covariates observed by the planner consistent with a transfer scheme $T(\tilde{\mathbf{X}}_i)$, $T(\tilde{\mathbf{X}}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, $\bar{\tau}^T$ be the average distortion from responses to the transfer function, and $\bar{g}$ be the average efficiency gain. If the average absolute targeting error exceeds twice the average efficiency gain minus the average funding gap; i.e.,*

$$\frac{1}{N}\sum_i |\tilde{c}_i - T(\tilde{\mathbf{X}}_i)| > 2\bar{g} - \bar{\Delta} - \bar{\tau}^T,$$

then there is no distribution of g with $g_i \geq 0 \ \forall i$ for which the policy and transfers create a Pareto improvement.

The bottom line of this mechanism design interpretation is that the impossibility condition is the same, but, all else equal, the challenge of creating a Pareto improvement is harder because the transfer scheme will create some distortion. This distortion can be incorporated into the impossibility condition by simply subtracting off the average distortion from the surplus.

Taking the condition to data: The mechanism design interpretation may pose an additional challenges in taking the condition to data. In some cases, a researcher may observe $\mathbf{X}_i$, rather than $\tilde{\mathbf{X}}_i$ (e.g., one sees covariates before a pollution tax and transfer system is implemented). In that case, it will not be possible to calculate the average absolute error. If, however, we believe (are willing to assume) that the observable covariates are no worse at predicting initial burdens than the adjusted covariates, then we can still proceed with the same empirical tests of the impossibility condition with data. This is idea is formalized in assumption 1:

Assumption 1. *A prediction of $\tilde{c}_i$ on $\tilde{\mathbf{X}}_i$ does not have a lower absolute average error than a prediction of $\tilde{c}_i$ on $\mathbf{X}_i$.*

The possibility of endogenous covariates also elevates the potential importance of interpreting the relationship between the transfer function $T(\cdot)$ and the welfare gains induced by the policy. One may wish to contemplate the use of covariates that are better predictors of initial burdens but are directly related to the externality (efficiency gain). For example, one might wonder about conditioning transfers for a gas tax based on vehicle ownership, or even mileage or fuel economy. Including these variables in the vector $\mathbf{X}$ would be expected to improve prediction, but they erode the incentives created by a pollution tax by effectively negating the incentives to reduce emissions via mechanisms that are compensated in the transfer function.

Conceptually, this is accounted for in the model already if we interpret G as being the efficiency gain that results from a combination of a policy and a particular transfer scheme. Intuitively, including these kinds of covariates would shrink prediction errors but it would also reduce the size of the welfare gains. Thus, the challenge for using the impossibility condition when such variables are included is not in the derivation of the condition, but in the calculation of the G that would be consistent with a given $T(\cdot)$. For the case of the gasoline tax, I show that the impossibility condition holds even after including a set of such regressors without assuming any reduction in G.

3 Consumption data on externality-generating goods

This paper uses data from the interview portion of the Consumer Expenditure Survey (CEX), which is a nationally representative sample of U.S. households, from 1996 to 2016. The CEX defines a unit of observation as a consumer unit, which is a set of individuals who reside together and are either related by blood or marriage, or who make financial decisions together.

Interviews consist of retrospective questions that ask about the consumer unit's total expenditures on various items over the prior three months. Units are interviewed four times, once each quarter, but not all units complete all four rounds of interviews. For the analysis below, expenditure categories are averaged over however many interviews are completed by a consumer unit, and then scaled to represent annual consumption amounts.

Table 1 shows summary statistics on expenditures. Key for this paper is that there is wide variability in the consumption of all variables. For example, average consumer unit expenditures on motor fuels is \$1,820, but the standard deviation is nearly as large as the mean, at \$1,716.

I have two chief concerns with using the CEX for this study. First, the analysis is concerned with variance and predictability of consumption levels across households. The survey response may mismeasure true consumption either because of sampling variability or because of inaccuracies in self-reported responses.[5] To address this concern, in appendix B, I show that key results are robust to the use of two other data sets that have better

[5] For a discussion of CEX data quality, see Meyer, Mok, and Sullivan (2015).

Table 1: Household Expenditure Statistics by Category

	Mean	Median	St. Dev	CV	Pct 0
Motor fuels	\$1,820	\$1,398	\$1,716	0.9	9%
Electricity	\$1,143	\$984	\$913	0.8	9%
Natural gas	\$413	\$162	\$611	1.5	42%
Alcohol	\$230	\$14	\$485	2.1	48%
Tobacco	\$318	\$0	\$788	2.5	71%
All energy	\$3,377	\$2,933	\$2,423	0.7	3%
All sin goods	\$3,925	\$3,411	\$2,757	0.7	2%

Table shows annualized expenditures by category for all households in sample (N=197,668). Dollar amounts are in \$2015. Statistics are weighted by survey sample weights. All energy sums motor fuels, electricity and natural gas. All sin goods includes all five individual categories summed. CV is the coefficient of variation. Pct 0 is the percentage of consumer units reporting zero expenditures in the category.

Table 2: Summary Statistics of Demographic Variables

	Mean	St. Dev.	Min	Max
Before-tax income (\$2015)	59,224	61,678	-419,200	971,100
Consumer unit (CU) size	2.4	1.5	1	29
Persons $<$ 18 in CU	0.62	1.1	0	14
Persons $>$64 in CU	0.28	0.6	0	8
Urban indicator	0.91	0.29	0	1
Reference person married	0.50	0.50	0	1
Year	2006	6.1	1996	2016

measures of gasoline consumption (the National Household Travel Survey) and home energy (the Residential Energy Consumption Survey), but are available for only one recent year.

Second, the CEX reports expenditures, not quantities. To model an ad valorem tax on a product, only the total expenditure is required. Corrective taxes, however, will often take the form of a specific (per unit) tax. For example, a carbon tax will raise the price of gasoline by a constant amount per gallon. Thus, to model the impact of a carbon tax on gasoline consumption, we need to estimate the gallons of gasoline consumed by a household, based on their reported expenditure and prices.

For gasoline and diesel fuels, I use data from the Energy Information Administration (EIA) on the sales-weighted, tax-inclusive, retail price of all grades of each fuel type at the closest available geographic match to the consumer unit. That is, where the CEX identifies a consumer unit's metropolitan statistical area and the EIA has city-specific prices, the consumer unit is assigned prices in the past quarter that are the average EIA price for that city. In other cases, matches must be made at the state or PADD level.

For other goods, determining the price paid by consumers is more challenging. Consider alcohol. Prices will vary widely if a consumer unit is purchasing low cost beer or high-end Scotch. As a result, for goods other than motor fuels, I focus on predicting expenditures directly (rather than predicted tax burdens), which translates directly to taxes under an ad valorem tax, recognizing that this is not how a true Pigouvian tax would be designed.

The core empirical task in the paper is to determine the degree to which demographic variables that might plausibly be used in a transfer function are able to predict variation in expenditures across consumer units. Table 2 summarizes the key variables used for this purposes, which are measures of income, household size and location.

4 Empirical tests of the impossibility condition

4.1 A gasoline tax creates losers

The primary empirical application of this paper is a gasoline tax. The conceptual goal of this analysis is to analyze an optimally designed Pigouvian tax. I thus focus on the gasoline tax as a well-targeted policy for correcting carbon externalities, but I discuss the implications of other driving-related externalities in the robustness section below.

In this section, I calculate the relative magnitude of welfare gains as compared to revenue raised from a motor fuel tax, and then demonstrate the degree to which demographic variables can predict motor fuel consumption. Specifically, I model a small tax increase of 10 cents on motor fuels (both gasoline and diesel) under the assumption that the carbon externality from motor fuel consumption is not corrected at all prior to the tax. That is, I am interpreting existing gasoline and diesel taxes as having been motivated by considerations about the optimal way to raise revenue, irrespective of a carbon externality. These assumptions are designed to be conservative against my findings, as they will maximize the implied welfare gains from carbon taxation.

4.1.1 What are the carbon externality gains from motor fuel taxation?

As described in the model, the welfare gain from a small tax on gasoline will be equal to the change in gasoline consumption induced by the tax times the externality per gallon. I assume that in the long run a gasoline tax will be borne completely by consumers so that prices will rise by 10 cents per gallon.[6]

The gasoline demand literature typically estimates elasticities, so I translate the 10 cent gasoline hike into a percentage price change using the average retail gasoline price facing the consumer unit at the time of the survey in its geographic location. I then use a gasoline price elasticity of -0.4, which is interpreted as a long-run price elasticity, to translate this price change into a change in gallons of fuel consumed.[7] By its very nature, it is challenging to estimate the long-run price elasticity of gasoline. I experiment with alternative values below.[8]

I use the EPA's conversion factor to determine the tons of carbon emitted per gallon of gasoline consumed (17.6 pounds per gallon / 2205 pounds per metric ton for E10, or 22.5 pounds per gallon / 2205 pounds per metric ton for diesel) and then multiply by $40 for the social cost of carbon, which is consistent with official government values in the last year of the sample data. This translates to an externality of $0.31 per gallon. Note that using this value to calculate the efficiency gains to current drivers assumes that all benefits from

[6] Existing studies find evidence of high pass through rates for state gasoline taxes, with many studies consistent with full pass through Chouinard and Perloff (2004, 2007); Doyle and Samphantharak (2008); Marion and Muehlegger (2011). Fewer studies consider the federal gas tax, perhaps because it has changed much less often, which impedes econometric investigation. Chouinard and Perloff (2004) conclude that only half of a federal tax increase is borne by consumers. If true, it would be important to consider the incidence on U.S. households through the producer side in interpreting the estimates.

[7] Small and Van Dender (2007) estimate long-run elasticities closer to half this magnitude. Hughes, Knittel, and Sperling (2008) conclude that the elasticity has been declining over time, finding preferred estimates well below -0.4. Espey (1998) finds a range of estimates that extend well beyond -0.4 in magnitude, but this is based on a variety of studies with varying credibility of empirical strategy. There is some suggestion that demand might respond more to gasoline taxes than price variation (Davis and Kilian 2011; Li, Linn, and Muehlegger 2014), though these estimates, taken from monthly changes in consumption, may be due inflated estimates due to consumers pre-buying in anticipation of price changes (Coglianese, Davis, Kilian, and Stock 2017). This difference seems unlikely to persist in the long run.

[8] I assume a homogeneous elasticity. Simple back of the envelope calculations make clear that allowing for heterogeneity will have unimportant impacts on the qualitative results because the tax is small.

Table 3: Summary of the Impact of a 10-cent Gasoline Tax

	Mean	Standard deviation
Annual gallons consumed	926	75
Price change	6%	4%
Change in gallons	-26	32
Initial burden (c)	$91	$75
Net revenue (r)	$90	$74
Externality gain (g)	$8.3	$10

Table summarizes the impact on private welfare, the externality and revenue of a 10 cent gasoline tax, assuming an elasticity of -0.4.

climate mitigation accrue to current drivers, which is an extreme interpretation designed to make a Pareto improvement more feasible. I discuss alternative values in section 4.1.5.

4.1.2 Externality gains are much smaller than the initial burden and revenue raised

Because I am modeling a small gasoline tax, the initial burden (loss of consumer surplus from the higher price) will be approximately equal to the revenue, both of which are simply the price increase times the number of gallons of gasoline consumed by the consumer unit. But, to be more precise, I use the elasticity estimate to calculate the final quantity consumed, and use that to calculate revenue. The welfare loss is calculated using a linear approximation. Specifically, revenue raised from each household is equal to 10 cents times the new consumption level, which is equal to the current observed level of consumption (from data) minus the elasticity (-0.4) times the implied change in price (current price plus 10 cents divided by the current price, all minus 1). The initial private welfare loss is calculated as the new consumption level (as described above), plus the triangle, which is the change in consumption (as described above) times $1/2$ times the tax (10 cents).

Table 3 shows these calculations for the estimation sample. The externality gains are $8.3 per consumer unit per year on average, while the revenue raised is $90 per consumer unit per year. Average costs imposed on consumers is slightly higher, at $91. The revenue raised is an order of magnitude larger than the externality gain. This has an important implication for the ability of the planner to create a Pareto improvement because, as shown by the theory, the externality gains represent the "error budget" available. A large amount of revenue needs to be reallocated via a transfer function, and the error budget is small relative to the revenue raised.

4.1.3 Most variation in burdens is not predictable

The key suggestion of the theoretical model is that the degree to which the initial (pre-transfer) burden of the corrective tax can be predicted by variables that are used in the transfer function will determine whether a Pareto improvement is technologically feasible. Simple regression of the household level burden on variables that constitute the transfer function thus provides the required estimates. Below, I present results where the left-hand side variable is the estimated household level initial burden of a 10 cent gas tax.[9] All values are inflation adjusted to 2015.

The theory involves non-squared errors, so I present least absolute deviation (LAD) regressions that will minimize non-squared errors. But, I also present parallel specifications from OLS because the properties of OLS and the R^2 goodness of fit statistic is most familiar. Note that LAD will, by definition, yield lower absolute errors, but OLS, by definition, will maximize the R^2.

Table 4 presents the primary estimates from this exercise, with the top panel reporting OLS results. All regressions include year of sample fixed effects, which account for any time trends, though it turns out that excluding them has almost no impact on the results. Designing a transfer scheme that depends on any variables that are not strictly exogenous will create distortionary incentives. As a result, I focus attention first on the "most exogenous" variables that are likely components of a tax scheme, which are demographic indicators for household structure and geographic indicators for state and urban versus rural. Specifically, regressions include state dummies, an urban indicator, and dummy variables for the number of people in the household, as well as the number of minors, and the number over age 60. These variables predict just under 30% of the variation in gasoline tax burdens.

Column B adds a linear income control, followed by a non-parametric function of income (dummies in five-year bins) in column C. These provide a modest boost in the explanatory power, with the R^2 bumping up to .331 and .356, respectively. For reference, income by itself, without any demographic or geographic variables, explains only about 15% of variation (results not shown). Column C is my preferred specification. It is based on characteristics that are already part of the tax system, and could plausibly be used to design a tax reform or transfer scheme that accompanies an externality-correcting tax.

The unexplained variation in this specification is far too large to achieve a Pareto improvement. The average absolute error allows for direct comparison with the welfare gains from the externality. The residuals are around \$45 per household. This compares to the \$8.25 welfare gain. This is directly related to Condition 1: as long as the absolute average

[9] Because I am assuming a homogenous elasticity across households, this is equivalent to using initial baseline consumption (in gallons) as the left-hand side variable.

Table 4: Predictability of Burden of a 10-cent Gasoline Tax

OLS	A	B	C	D
Avg. Abs. Error	\$46.6	\$45.0	\$44.2	\$39.9
R^2	.292	.331	.356	.456
LAD	E	F	G	H
Avg. Abs. Error	\$45.7	\$44.1	\$43.2	\$38.8
Pseudo-R^2	.181	.210	.226	.306
N	197,668	197,668	197,668	197,668
Year FE	Y	Y	Y	Y
Demo & geo controls	Y	Y	Y	Y
Linear income		Y	Y	Y
Binned income			Y	Y
Vehicles & energy				Y

Each letter represents a unique regression predicting the initial burden from a 10 cent gasoline tax. A and E include year fixed effects and dummy variables for number of household members, reference person married, number in household over 64, number under 18. B and F add a linear control for before tax household income. C and G add dummies for every \$5,000 of income. D and H add dummies for the number of vehicles owned or leased and level variables of expenditures on natural gas, electricity and heating oil.

error exceeds twice the welfare gain, a Pareto improvement is not possible. Moreover, it is not just a matter of a few people being left as net losers. The best fitted scheme leaves more than one-third of households as net losers, even with the generous assumptions employed throughout.

Column D adds some clearly endogenous variables that would create significant distortions and are thus likely problematic variables for inclusion in a transfer scheme, including home energy consumption and dummies for the number of vehicles owned by the household, and dummies for the number leased. These variables do provide an additional boost to explanatory power, but even with vehicle ownership variables included, the variables explain less than half of the variation.

The bottom panel of table 4 shows LAD specifications. As expected, these lower the absolute error for identical specifications, but only by a very small amount.

Figure 1 shows the distribution of net losses, accounting for both the externality gain and the targeted transfers, based on column C in Table 4. A full 37% of households remain

Figure 1: Net Loss from 10-cent Gasoline Tax with Targeted Transfer

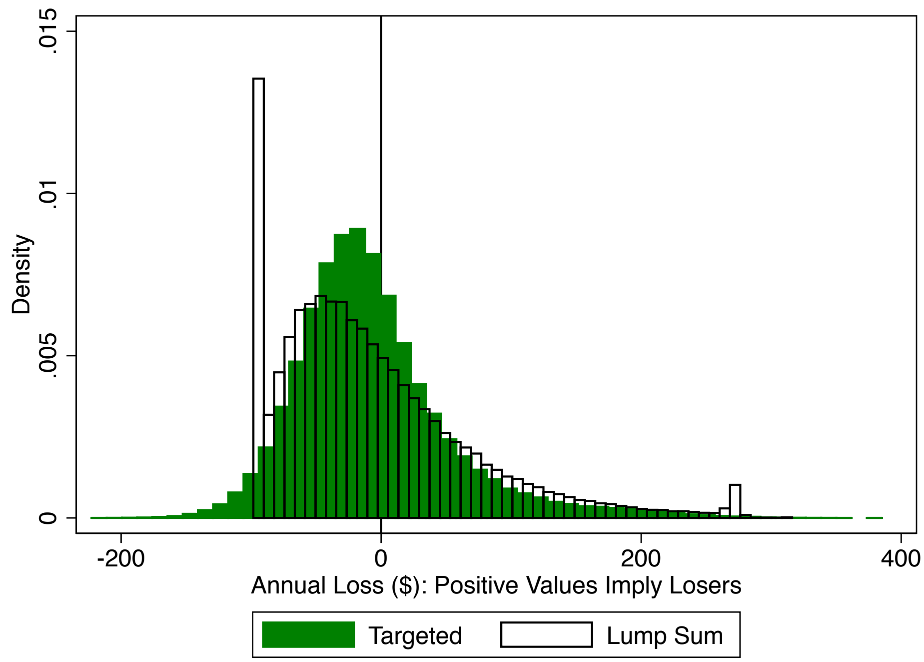


Figure shows the distribution of net impacts of a 10-cent gasoline tax, in dollars per year. A positive value implies a welfare loss. Results for equal per household transfer in transparent. Solid green indicates results for transfer scheme based on specification C in Table 4. The net impact is the private welfare loss, net of the targeted transfer scheme, net of the externality gain, which is assumed to be equal to each household.

as net losers under this scheme.

For comparison, the figure also shows the distribution of net losses under a scheme where all households are rebated an equal share of the revenue. A similar fraction of households are net losers under both of these scenarios, but targeting radically reshapes the distribution.

The variables chosen here are the ones that are most likely to be used for a transfer scheme that operates through the tax code. The tax code is essentially a function of income and demographic structure of the household. As such, I interpret the results of Table 4 as demonstrating that gasoline expenditures are not predicted well enough to come remotely close to enabling a Pareto improvement. A Pareto improvement is not feasible.

It is worth restating the nature of the prediction dilemma at this point. The challenge here is not that the CEX is a sample needed to predict out of sample consumption. The problem pointed out here is not solved by having a census of burdens for all households. Given that census, the planner could simply rebate every household exactly the burden imposed on it. But, if households understand this, then it completely (or at least significantly) undoes the price incentive—gasoline is not more expensive because the tax increase is rebated back, so

Table 5: Lasso Regressions on Burden of 10-cent Gasoline Tax

	OLS (C)	Lasso	Lasso
Avg. Abs. Error	\$44.16	\$44.23	\$43.16
R^2	.356	.353	.379
Vars. Supplied		166	3,352
Vars. Selected		135	1,855
N	197,668	197,668	197,668
Year FE	Y	Y	Y
Demog. & geog. controls	Y	Y	Y
Linear income	Y	Y	Y
Binned income	Y	Y	Y
Additional interactions			Y

The first column repeats the OLS regression from Column C of Table 4. The second column runs a lasso regression on the same right hand side variable to perform a check for overfitting. The third column runs lasso with a large set of additional interactions. See text for details.

there will be no externality gain. In terms of the model, this is the case of the endogenous $\mathbf{X}$ vector, where the transfer $T(\cdot)$ would shrink efficiency gains g_i down to zero.

The impossibility conditions in the theory section are derived for a generic transfer function. The regressions here are designed to find a transfer function that minimizes the average absolute error. The regressions show the smallest attainable average absolute error, given the allowed vector of covariates. If these errors are still so large as to make the impossibility condition hold, then we conclude that all feasible transfer functions would trigger the impossibility condition.

4.1.4 Machine learning marginally improves prediction

The problem posited here is fundamentally a prediction problem. It is thus a natural application for machine learning. A simple version is pursued here to see if initial steps can dramatically improve prediction.

Table 5 reports results of lasso regressions that predict the variation in tax burdens. The first column repeats column C from Table 4 for reference. The second column reports results from a lasso regression on the same variables to check for overfitting in the main specification. The specification uses a 10-fold cross validation and experiments with a range of lasso penalty parameters. Results suggest minimal overfitting. Lasso chooses a zero coefficient on 29 out of 166 variables, but this results in economically insignificant changes to prediction accuracy.

The third column introduces several thousand additional variables and uses the same 10-fold cross validation to select variables for inclusion, with the lasso penalty parameter chosen endogenously by the optimizer. Because the main specification includes predominantly binary dummy variables, the focus is on interactions, rather than higher order polynomials. The third column includes interactions of every income category and income linearly with year, state dummies, urban indicator, family size dummies, number in household under 18 dummies, number in household over 64 dummies, and a dummy for marital status. State by year fixed effects are also included. Despite selecting over 1,800 variables for inclusion, the improvement in prediction is minimal, and, from the point of view of achieving a Pareto improvement, barely perceptible. Additional experimentation with other interactions of these core variables produced similar results.

This is only a basic attempt to introduce prediction methods, but the lack of significant improvement from broader specification searches suggests that the variation in the burdens in the CEX is not predictable with the set of cross-sectional measures (household demographics, location of residence, and income) that is most plausibly usable as part of the tax code.

4.1.5 Robustness to parameter choices

In this section, I present results that alter three assumptions about the data. First and most simply, I increase the number of data points that are winsorized. Second, I modify the elasticity of gasoline consumption from -0.4 to -0.6 and then -0.8, to reflect higher estimates from the literature. Greater elasticities are important because they will lead to greater welfare gains, which aids the elimination of losers. Third, I greatly increase the externality per gallon of gasoline consumed, from around \$0.31 to \$2.

For a given level of prediction accuracy and revenue gap, it is straightforward to calculate the mean efficiency gain that would cause the impossibility condition to hold exactly. In this case, it can be found by taking the average absolute error from Table 4, Column C, (\$43.2) and adding the average revenue gap (\$1), and dividing by the number of gallons saved on average from the tax (26.7). This yields a mean efficiency gain of \$1.66 per gallon. That is the externality that would cause the impossibility condition to hold exactly.

Such a high number could be justified based on either of two approaches. First, the social cost of carbon is a highly uncertain number. A prominent recent study suggested a central estimate of \$185 per metric ton, which translates to \$1.43 per gallon for E10 (Rennert et al. 2022). Interpreting this full value as accruing to current drivers is still implausibly generous to the case, but it is meant to illustrate the possible importance of a much higher social cost of carbon.

Another argument for higher welfare gain might come from considering other externali-

Table 6: Fraction of Losers Under Alternative Assumptions

Elasticity	Externality per gallon	Percent Winsorized	Percent Losers
-0.4	\$0.31	1%	37.0%
-0.4	\$2	1%	15.5%
-0.6	\$2	1%	9.7%
-0.8	\$2	1%	6.2%
-0.8	\$2	10%	3.0%

Each row comes from a separate regression of the burden of a 10-cent gasoline tax on the same set of covariates as specification C in Table 4. Each row varies a parameter as listed in the first three columns.

ties. Harrington, Parry, and Walls (2007) survey the literature and conclude that greenhouse gas emissions externalities are quite modest compared to accident and congestion externalities (though this was at a much lower social cost of carbon than is now preferred by many). A gas tax is a very poor instrument for targeting congestion, and a mediocre at best instrument for targeting accidents or local air pollution. Nevertheless, I now show cases where the gas tax could have much larger benefits in order to compare results.

In arriving at a \$2 per gallon externality, I modify the values from Harrington, Parry, and Walls (2007) to account for a higher accident externality, at \$0.91 per gallon based on Anderson and Auffhammer (2014), but interpret the carbon benefits as negligible. I then subtract off the sales-weighted average gas tax in the US of \$0.48. In terms of the literature on second-best gasoline taxes, however, note that this is still a generous interpretation in that it ignores fiscal interactions that exacerbate labor market distortions. Parry and Small (2005), for instance, argue that the second-best tax is only around 60% of marginal damages due to fiscal interactions.

At these higher values, the impossibility condition no longer binds, but that is necessary, not sufficient, for creating a Pareto improvement. Table 6 uses targeted transfers from specification C from Table 4, under alternative assumptions, to calculate the number of households that are net losers. Dramatically increasing the interpreted externality gain per mile roughly halves the number of households who are net losers from a gasoline tax. Increase the elasticity of gasoline consumption to much higher rates further drives down the fraction of losers. In this scenario, the number of net losers is driven down to 6%. This is a modest number, but it should be kept in mind that there are many generous assumptions deployed in this case, so it should be interpreted as a frontier possibility rather than a realistic point estimate. Even in this case, some households are net losers. Finally, taking all of the prior assumptions and also winsorizing a full 10% of the data drives down the number of losers to 3%. This suggests the potential for sharply limiting the number of losers, but only when a variety of optimistic cases hold simultaneously.

4.2 Burdens from other externality-correcting taxes are no easier to predict

The focus of this paper empirically is on a gasoline tax, but the CEX enables me to make quick assessment of the degree of predictability of other consumption categories that might be the focus of sin taxes. A gas tax has the advantage that it is relatively easy to translate expenditure data into quantities using gasoline price information, and hence to estimate the impact of a specific (per gallon) gasoline tax. The impact of other sin taxes is more difficult to determine because the goods are more heterogenous (e.g., there are many types of alcohol) and are subject to non-linear prices (e.g., two-part tariffs for electricity and natural gas).

Nevertheless, a broad picture of heterogeneity and predictability can be gained by simply regressing total expenditures in these categories on the demographic variables to see how much of the baseline expenditure variation is predictable. This exercise would exactly mimic the burden of an ad valorem sin tax, and they likely come close to mimicking the scale effect of sin tax levied per unit of the sin good in question.

Note that Table 1 shows that electricity has a similar coefficient of variation with motor fuels, but that other categories have even larger variability. OLS regressions in Table 7 shows the same pattern in terms of predictability. Electricity consumption is very similar in its predictability to motor fuels, but other sin goods are substantially harder to predict.

This analysis is incomplete, as it does not account for the welfare gains and is based on an ad hoc assumption about how a corrective tax would impact prices. But, the results suggest that a gasoline tax is likely the easiest place to achieve broad gains, and that the other externality-creating goods are likely to create even larger numbers of losers because of the greater inability to predict variation in baseline expenditures.

4.3 Historical baselines

In a cross section, the consumption of externality-creating goods is difficult to predict. In a panel, there may be more scope for prediction if one can base the transfer function on past consumption or behavior. Moreover, this can be efficient *if* agents are unaware of the fact that behavior in a time period will be used to determine future transfers until after the fact. Indeed, transfers based on historical emissions are often used in cap and trade systems,

Table 7: Predictability of Other Sin Expenditures (OLS)

All statistics are R^2	A	B	C
Motor Fuels	.336	.382	.403
Electricity	.281	.324	.327
Natural gas	.179	.211	.214
Alcohol	.051	.126	.129
Tobacco	.043	.046	.050
All energy	.399	.471	.490
All sin goods	.367	.441	.459
N	197,668	197,668	197,668
Year FE	Y	Y	Y
Demog. & geog. controls	Y	Y	Y
Linear income		Y	Y
Binned income			Y

Each entry in the table is the R^2 from a separate regression that predicts expenditures (not burdens) on the category listed in the row, with control variables that vary by column. Column A includes year fixed effects and dummy variables for number of household members, reference person married, number in household over 64, number under 18. Column B adds a linear control for before tax household income. Column C adds dummies for every \$5,000 of income.

which frequently give away permits as a function of emissions in a baseline period before policy was implemented (Schmalensee and Stavins 2017).

This is harder to envision for households, but not impossible. The challenge is that policymakers do not already have a census of consumption behavior for most relevant goods. New data collection systems would need to be created to, for example, measure every household's consumption of gasoline. The challenge would be to create new systems without making actors believe that the systems will be used to determine future transfers, which is important lest actors strategically manipulate baseline measures. (For cap and trade systems, pollutants were already regulated and measured in most systems, which mitigated this concern.) These considerable practical challenges notwithstanding, it is worth asking how well past behavior predicts future consumption for relevant goods. Unfortunately, the CEX is a poor data source for studying panel variation, but other data sources can give some guidance on how past consumption data might be used.

Sallee and Tarduno (2022) study an experimental congestion pricing scheme in the Seattle metropolitan area. A group of 255 volunteers were enrolled in the study and had a transponder installed in their vehicle, but the nature of the future pricing experiment was not explained at that time. Driving patterns were measured for several months. Then, a road pricing scheme was implemented where drivers had to pay a surcharge to use key roads during peak hours. Surveys taken throughout the process provided demographic covariates.

Sallee and Tarduno (2022) use the observed congestion charges borne by drivers, taking their behavioral adjustments into account, as a measure of $\tilde{c}_i$ and show (a) household demographics predict a modest fraction of cost variation ($R^2 = .24$); (b) neighborhood demographics explain very little variation ($R^2 = .04$); and (c) baseline "tolls" (the tolls that a household would have paid had the pricing program been in affect during the baseline driving period) explains much more ($R^2 = .57$). This suggests that baseline consumption is a better predictor than other covariates. But, Sallee and Tarduno (2022) find that the impossibility condition still holds, with average absolute errors being more than three and a half times the estimated average efficiency gain in this case.

Panel data on California electricity: In this section, I discuss what may be the most optimistic case for the use of panel data, which is home energy consumption. This is an optimistic case because comprehensive measures of consumption do in fact exist, though they reside in a patchwork of utilities across the country and are not centralized in a government agency.

For the analysis, I gained access to a large panel of electricity billing data from the universe of households that are served by California's three large investor owned utilities, Pacific Gas and Electric (PG&E), Southern California Edison (SCE), and San Diego Gas

Table 8: Electricity consumption prediction using historical data

Dependent variable = annual kWh consumed in 2019	PG&E	PG&E	SDG&E	SDG&E	SCE	SCE
2018 annual kWh	0.99	1.02	0.92	0.93	0.93	0.94
	(0.004)	(0.018)	(0.004)	(0.021)	(0.004)	(0.023)
Avg. Abs. Error (kWh)	851.27	852.88	1053.29	1054.15	937.53	934.31
R^2	0.97	0.97	0.96	0.96	0.76	0.76
N	3,093,970	3,093,970	1,254,364	1,254,364	4,304,757	4,304,757
Unit observed	account	account	account	account	meter	meter
2017 annual kWh	N	Y	N	Y	N	Y
Rooftop solar	N	Y	N	Y	N	Y
CARE participation	N	Y	N	Y	N	Y

Table reports OLS regression coefficients and related statistics for regressions from three different utilities in California. The dependent variable is annual consumption in 2019 (in kWh), which is regressed on an intercept and the prior year's consumption. Some specifications include 2017 consumption, an indicator for whether a household has rooftop solar or participates in a low-income pricing program. Data for PG&E and SDG&E are at the level of an account, but SCE data are linked by meter.

and Electric (SDG&E). I have account-specific monthly billing data that includes electricity consumption monthly from 2017 to 2019. For this analysis, I take annual average consumption for all households that have complete records over all three years, which is more than 8 and a half million accounts.[10] For PG&E and SDG&E, the records indicate a unique customer account attached to a particular residence. For SCE, the data indicate a unique identified for a particular meter, so in the panel data there is some variation due to changes in the account holder at an address that I cannot observe.

Table 8 reports the R^2 and average absolute error from regressions that use 2018 consumption to predict 2019 consumption across households. The coefficients are close to one, and the R^2 are extremely high.[11] In some specifications, I also include 2017 annual consumption and dummy variables for participation in a low-income pricing program (CARE) and whether a household has rooftop solar. Adding these has minimal impact. Overall, these results show that energy consumption is quite stable, so historical data can lead to a dramatic improvement in prediction quality as compared to using only covariates.

[10] The billing data have no demographic information or other covariates, except location. For that reason, I do not use the billing data for the preceding analysis, but do use it here just to test the potential of using historical consumption.

[11] The R^2 for SCE is much lower. This is probably because different households are moving in and out of a particular location, so much of the unexplained variation is likely due to the fact that the same meter is linked to different households over time.

However, even in this case the impossibility condition holds depending on parameter choices. For ease of interpretation, consider a 1 cent per kWh tax. The average absolute error for each utility can then be calculated by just multiplying the values in table 8 by $0.01 to get an estimate of $\tilde{c}_i$. For example, this is $8.51 for PG&E.

The welfare gains from such a tax depend on the elasticity, the current price, and the externality. I use an elasticity of -0.5, which is almost certainly too high, and I take current prices in PG&E ($0.25 per kWh) and the externality (0.39 tons of carbon dioxide equivalent per MWh) from Borenstein, Fowlie, and Sallee (2021). The price and elasticity values imply that a 1 cent increase in price creates a 4% change in price and a 2% change in quantity, equivalent to 121 kWh in PG&E, where average consumption is 6,031 kWh per year. At a $40 per ton social cost of carbon, 121 kWh (equates to 0.047 tons) is a per household reduction in the externality of only $1.88.[12] Two times this value is far below the average absolute error, so the impossibility condition continues to hold, even where the predictability of consumption is so strong. The impossibility condition breaks when the social cost of carbon is around $100 a ton. This is a plausible value, but the exercise demonstrates that, even where predicability is highest, the impossibility condition still holds for plausible parameter values.

I interpret this exercise as showing that, on the one hand, the use of historical consumption can radically improve predictability, but that this alone is not sufficient to ensure that the impossibility condition does not hold. Thus, I interpret the billing data analysis as suggesting yet again that a Pareto improvement is implausible, if not impossible.[13]

5 Conclusion

This paper uses theory and data to argue that policies like Pigouvian taxes—which improve social efficiency but create heterogeneous costs and benefits—will inevitably create some losers because transfers targeting the losers will tend to be imprecise.

The theory demonstrates how one's ability to compensate losers depends on the predictability of heterogeneous policy burdens and the size of efficiency gains. The theory delivers a specific test that can be taken directly to data. Empirically, the case of a gasoline

[12] The marginal price of electricity is actually above the social marginal cost in these utilities because California utilities recover system costs via high volumetric (marginal) prices (Borenstein, Fowlie, and Sallee 2021). As a result, efficiency requires that we lower prices, not increase them. The exploration here should be interpreted with this in mind.

[13] These results are driven in large part by the fact that there is much greater heterogeneity in measured kWh consumption in these billing data than is implied in the equivalent data sets from the CEX or RECS. In the billing data, the coefficient of variation in annual consumption in all three utilities is around 2, whereas that value is close to 1 in the CEX. This large variation means that, even with a very high R^2, the average absolute errors turn out to be large.

tax is considered, and the possibility of a Pareto improvement is soundly rejected. Preliminary evidence on other externality creating goods suggests the same conclusion. In short, Pigouvian taxes create losers.

This is an important conclusion as it suggests the need for nuance in a range of important policy debates. Economists sometimes argue that efficiency-enhancing policies, at least in principle, can be paired with targeted transfers so as to rationalize completely abstracting from distributional implications and judging policies purely on efficiency grounds. This paper argues for more caution in this line of reasoning. The fact that a policy creates losers is not in and of itself a reason to reject the policy, but it does point out one reason why efficiency enhancing policies may not prevail in the policy-making process and more broadly suggests a shift in our approach to discussing the tension between equity and efficiency.

References

Akerlof, George A. 1978. "The Economics of 'Tagging' as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning." *American Economic Review* 68 (1):8–19.

Anderson, Michael and Maximilian Auffhammer. 2014. "Pounds that Kill: The External Costs of Vehicle Weight." *Review of Economic Studies* 82 (2):535–571.

Atkinson, Anthony B. and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6 (1-2):55–75.

Borenstein, Severin, Meredith Fowlie, and James M. Sallee. 2021. "Designing Electricity Rates for An Equitable Energy Transition." Energy Institute Working Paper 314.

Bovenberg, A. Lans and Lawrence H. Goulder. 2001. "Neutralizing the Adverse Industry Impacts of CO2 Abatement Policies: What Does it Cost?" In *Behavioral and Distributional Effects of Environmental Policy*, edited by Carlo Carraro and Gilbert E. Metcalf. NBER, 45–90.

Bovenberg, A. Lans, Lawrence H. Goulder, and Derek J. Gurney. 2005. "Efficiency Costs of Meeting Industry-Distributional Constraints Under Environmental Permits and Taxes." *RAND Journal of Economics* 36 (4):951–971.

Bovenberg, A. Lans and F. van der Ploeg. 1994. "Environmental policy, public finance and the labour market in a second-best world." *Journal of Public Economics* 55:349–390.

Burtraw, Dallas and Karen Palmer. 2008. "Compensation Rules for Climate Policy in the Electricity Sector." *Journal of Policy Analysis and Management* 27 (4):819–847.

Burtraw, Dallas, Richard Sweeney, and Margaret Walls. 2008. "The Incidence of US Climate Policy: Where You Stand Depends on Where You Sit." Resources for the Future Discussion Paper.

Chouinard, Hayley and Jeffrery M. Perloff. 2004. "Incidence of Federal and State Gasoline Taxes." *Economics Letters* 83 (1):55–60.

Chouinard, Hayley H. and Jeffrery M. Perloff. 2007. "Gasoline Price Differences: Taxes, Pollution Regulations, Mergers, Market Power, and Market Conditions." *The B.E. Journal of Economic Analysis & Policy* 7 (1 (Contributions)):1–26.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. *Targeting of Transfers in Developing Countries: Review of Lessons and Experience*. Washington, DC: World Bank.

Coglianese, John, Lucas W. Davis, Lutz Kilian, and James Stock. 2017. "Anticipation, Tax Avoidance, and the Price Elasticity of Demand for Gasoline." *Journal of Applied Econometrics* 32 (1):1–15.

Cremer, Helmuth, Firouz Gahvari, and Norbert Ladoux. 1998. "Externalities and Optimal Taxation." *Journal of Public Economics* 70:343–364.

———. 2003. "Environmental Taxes with Heterogenous Consumers: An Application to Energy Consumption in France." *Journal of Public Economics* 87 (12):2791–2815.

Cronin, Julie Anne, Don Fullerton, and Steven E. Sexton. 2019. "Vertical and Horizontal Redistribution from a Carbon Tax and Rebate." *Journal of the Association of Environmental and Resource Economists* .

Davis, Lucas W. and Lutz Kilian. 2011. "Estimating the Effect of a Gasoline Tax on Carbon Emissions." *Journal of Applied Econometrics* 26:1187–1214.

Davis, Lucas W. and Christopher R. Knittel. 2019. "Are Fuel Economy Standards Regressive?" *Journal of the Association of Environmental and Resource Economists* .

Dinan, Terry. 2012. "Offsetting a Carbon Tax's Costs on Low-Income Households." Https://www.cbo.gov/publication/43713.

Doyle, Jr., Joseph J. and Krislert Samphantharak. 2008. "$2.00 Gas! Studying the Effects of a Gas Tax Moratorium." *Journal of Public Economics* 92 (3-4):869–884.

Espey, Molly. 1998. "Gasoline Demand Revisited: An International Meta-Analysis of Elasticities." *Energy Economics* 20 (3):273–295.

Fischer, Carolyn and Wiliam A. Pizer. 2019. "Equity versus Efficiency in Energy Regulation." *Journal of the Association of Environmental and Resource Economists* .

Gauthier, Stèphane and Guy Laroque. 2009. "Separability and Public Finance." *Journal of Public Economics* 93:1168–1174.

Goulder, Lawrence H., Marc A.C. Hafstead, and Michael Dworsky. 2010. "Impacts of Alternative Emissions Allowance Allocation Methods Under a Federal Cap-and-Trade Program." *Journal of Environmental Economics and Management* 60 (3):161–181.

Grainger, Corbett A. and Charles D. Kolstad. 2010. "Who Pays a Price on Carbon?" *Environmental and Resource Economics* 46 (3):359–376.

Harrington, Winston, Ian Parry, and Margaret Walls. 2007. "Automobile Externalities and Policies." *Journal of Economic Literature* 45 (2):373–399.

Hassett, Kevin A., Aparna Mathur, and Gilbert E. Metcalf. 2009. "The Incidence of a U.S. Carbon Tax: A Lifetime and Regional Analysis." *Energy Journal* 30 (2):155–178.

Hughes, Jonathan E., Christopher R. Knittel, and Daniel Sperling. 2008. "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *The Energy Journal* 29 (1).

Jacobs, Bas and Ruud A. de Mooij. 2015. "Pigou Meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-Best Pigouvian Tax." *Journal of Environmental Economics and Management* 71:90–108.

Kaplow, Louis. 1989. "Horizontal Equity: Measures in Search of a Principle." *National Tax Journal* 42 (2):139–154.

———. 2004. "On the (Ir)Relevance of Distribution and Labor Supply Distortion to Government Policy." *Journal of Economic Perspectives* 18 (4):159–175.

———. 2012. "Optimal Control of Externalities in the Presence of Income Taxation." *International Economic Review* 53 (2):487–509.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2014. "Gasoline Taxes and Consumer Behavior." *American Economic Journal: Economic Policy* 6 (4):302–342.

Marion, Justin and Erich Muehlegger. 2011. "Fuel Tax Incidence and Supply Conditions." *Journal of Public Economics* 95 (9-10):1202–1212.

Mathur, Aparna and Adele C. Morris. 2014. "Distributional Effects of a Carbon Tax in Broader US Fiscal Reform." *Energy Policy* 66:326–334.

Metcalf, Gilbert E. 2009. "A Distributional Analysis of Green Tax Reforms." *National Tax Journal* 52 (4):655–681.

———. 2018. *Paying for Pollution: Why a Carbon Tax is Good for America.* Oxford University Press.

Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29 (4):199–226.

Mirrlees, James A. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38 (2):175–208.

Musgrave, Richard. 1959. *Theory of Public Finance; A Study in Public Economy.* New York: McGraw-Hill.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Harvard University Press.

———. 1982. *The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities*. Yale University Press.

Parry, Ian W.H. and Kenneth A. Small. 2005. "Does Britain or the United States Have the Right Gasoline Tax?" *American Economic Review* 95 (4):1276–1289.

Pigou, Arthur C. 1932. *The Economics of Welfare*. London: Macmillan and Co., 4th ed.

Pizer, Wiliam A. and Steven E. Sexton. 2019. "Distributional Impacts of Energy Taxes." *Review of Environmental Economics and Policy* 13 (1):104–123. Manuscript: Duke University.

Poterba, James M. 1991. "Is the Gasoline Tax Regressive?" *Tax Policy and the Economy* 5:145–164.

Rausch, Sebastian, Gilbert E. Metcalf, and John M. Reilly. 2011. "Distributional Impacts of Carbon Pricing: A General Equilibrium Approach with Micro-Data for Households." *Energy Economics* 33:S20–S33.

Rennert, Kevin, Frank Errickson, Brian C. Prest, Lisa Rennels, Richard G. Newell, William Pizer, Cora Kingdon, Jordan Wingenroth, Roger Cooke, Bryan Parthum, David Smith, Kevin Cromar, Delavane Diaz, Frances C. Moore, Ulrich K. Müller, Richard J. Plevin, Adrian E. Raftery, Hana Sevcikova, Hannah Sheets, James H. Stock, Tammy Tan, Mark Watson, Tony E. Wong, and David Anthoff. 2022. "Comprehensive Evidence Implies a Higher Social Cost of CO_2." *Nature* .

Saez, Emmanuel. 2002. "The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogenous Tastes." *Journal of Public Economics* 83 (2):217–230.

Sallee, James M. and Matthew Tarduno. 2022. "Can Rebates Foster Equity in Congestion Pricing Programs?" UC-ITS-2020-26.

Sandmo, Agnar. 1975. "Optimal Taxation in the Presence of Externalities." *Swedish Journal of Economics* 77 (1):86–98.

Schmalensee, Richard and Robert N. Stavins. 2017. "Lessons Learned from Three Decades of Experience with Cap and Trade." *Review of Environmental Economics and Policy* 11 (1):59–79.

Small, Kenneth A. and Kurt Van Dender. 2007. "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect." *Energy Journal* 28 (1):25–51.

West, Sarah E. 2004. "Distributional Effects of Alternative Vehicle Pollution Control Policies." *Journal of Public Economics* 88 (3-4):735–757.

Williams, Roberton C., III, Hal Gordon, Dallas Burtraw, Jared C. Carbone, and Richard D. Morgenstern. 2015. "The Initial Incidence of a Carbon Tax Across Income Groups." *National Tax Journal* 68 (1).

A Appendix: Proofs

Condition 1. *Let $\tilde{c}_i$ be the initial costs from a policy, N be the number of agents, $\mathbf{X}_i$ be a vector of exogenous covariates observed by the planner, $T(\mathbf{X}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, and $\bar{g}$ be the average efficiency gain. If the average absolute targeting error exceeds twice the average efficiency gain minus the average funding gap; i.e.,*

$$\frac{1}{N}\sum_i |\tilde{c}_i - T(\mathbf{X}_i)| > 2\bar{g} - \bar{\Delta},$$

then there is no distribution of g with $g_i \geq 0\ \forall i$ for which the policy and transfers create a Pareto improvement.

The proof proceeds by stating the total size of all losses (initial burdens net of transfer, excluding the surplus gains) among losers as an algebraic expression of the average absolute error and the average funding gap. If the total losses among losers exceeds the total surplus gains enjoyed by participants in the market, then it is not possible that these surplus gains are distributed so as to compensate all losers.

For any transfer regime $T(\mathbf{X}_i)$ and distribution of costs $\tilde{c}_i$, partition the data into losers (anyone with $\tilde{c}_i - T(\mathbf{X}_i) > 0$) and winners (anyone with $\tilde{c}_i - T(\mathbf{X}_i) \leq 0$). Denote the set of losers as $i \in L$, with their number being N_L. Denote the set of winners as $i \in W$, with their number being N_W.

Having partitioned the data into winners and losers, one can define the total losses among losers (which will be positive by construction), denoted Z_L and the total losses (which will be negative by construction) among winners as Z_W:

$$Z_L \equiv \sum_{i\in L}(\tilde{c}_i - T(\mathbf{X}_i)) > 0 \qquad Z_W \equiv \sum_{i\in L}(\tilde{c}_i - T(\mathbf{X}_i)) < 0.$$

The sum Z_L is the total amount of loss among losers. If this loss exceeds the total efficiency gains $\sum_i g_i$, then it is impossible to achieve a Pareto improvement, because even if those gains are distributed in the most favorable way among all losers, there are not enough gains to compensate all losers. The goal now is to redefine Z_L in terms of the average absolute error among all i (because this relates to an empirically estimable object) and the average funding gap.

The funding gap is the amount by which initial costs exceed revenue available for transfers, $\Delta = R - C$, with the budget constraint implying that $\sum_i T(\mathbf{X}_i) = R$. The funding gap is equal to the sum of Z_L and Z_W: $\Delta = Z_W + Z_L$. (To see this: $\Delta = C - R = \sum_i(\tilde{c}_i - T(\mathbf{X}_i)) = Z_L + Z_W$, with the latter equality true because Z_L and Z_W just partition the full set.) We will want to relate absolute values and will want an expression to substitute out Z_W. Because Z_W is always negative and Z_L is always positive:

$$|Z_W| = |Z_L| - \Delta \tag{2}$$

For notational convenience, denote the average absolute error as $|\tilde{c}_i - T(\mathbf{X}_i)| = |\varepsilon|$. Denote the average absolute error among all individuals by $|\bar{\epsilon}|$, with $|\bar{\epsilon_L}|$ and $|\bar{\epsilon_W}|$ representing the average absolute error of losers and winners respectively. The average absolute errors are equal to Z divided by N for each group:

$$|\bar{\epsilon_L}| = \sum_{i \in L} |\tilde{c}_i - T(\mathbf{X}_i)| = \frac{|Z_L|}{N_L} \quad \text{and} \quad |\bar{\epsilon_W}| = \sum_{i \in W} |\tilde{c}_i - T(\mathbf{X}_i)| = \frac{|Z_W|}{N_W}. \tag{3}$$

The average absolute value of all of the data is simply the sample-size weighted average of the absolute average among winners and losers, which is written in terms of Z_W and Z_L via substitution of (5):

$$|\bar{\epsilon}| = \frac{N_L|\bar{\epsilon_L}| + N_W|\bar{\epsilon_W}|}{N_L + N_W} = \frac{|Z_L| + |Z_W|}{N}.$$

Substituting for $|Z_W|$ using (2) yields:

$$|\bar{\epsilon}| = \frac{|Z_L| + |Z_W|}{N} = \frac{|Z_L| + |Z_L| - \Delta}{N}.$$

Rearrange to solve for an expression of $|Z_L|$, which is positive by construction:

$$Z_L = |Z_L| = \frac{N}{2}|\bar{\epsilon}| + \frac{\Delta}{2}. \tag{4}$$

This is an expression for the total loss among losers. If this loss exceeds the sum of efficiency gains, then a Pareto improvement is not possible. I.e., a Pareto improvement is not possible if $Z_L = \frac{N}{2}|\bar{\epsilon}| + \frac{\Delta}{2} > \sum_i g_i$. Rearranging yields the result:

$$\frac{N}{2}|\bar{\epsilon}| + \frac{\Delta}{2} > \sum_i g_i \Leftrightarrow |\bar{\epsilon}| > 2\bar{g} - \bar{\Delta}.$$

The statement in the proof uses the definition of $|\bar{\epsilon}|$. □

Condition 2. *Let c_i be the initial costs from a policy assuming no behavioral responses, N be the number of agents, $\mathbf{X}_i$ be a vector of exogenous covariates observed by the planner, $T(\mathbf{X}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, $\bar{g}$ be the average efficiency gain, and $\bar{b}$ be the average private welfare gain from behavioral responses. If the average absolute targeting error exceeds twice the sum of the average efficiency gain and behavioral adjustment gain minus the average funding gap; i.e.,*

$$\frac{1}{N}\sum_i |c_i - T(\mathbf{X}_i)| > 2(\bar{g} + \bar{b}) - \bar{\Delta},$$

then there is no distribution of g with $g_i \geq 0$ $\forall i$ for which the policy and transfers create a Pareto improvement.

The derivation of this result is the same as above. The only difference is that losers and winners are identified and partitioned according to c_i (rather than $\tilde{c}_i$). The welfare gains, b_i are then added to the efficiency gains g_i.

The losses among losers and gains among winners is defined as:

$$|\bar{\epsilon_L}| = \sum_{i \in L} |c_i - T(\mathbf{X}_i)| = \frac{|Z_L|}{N_L} \text{ and } |\bar{\epsilon_W}| = \sum_{i \in W} |c_i - T(\mathbf{X}_i)| = \frac{|Z_W|}{N_W}. \tag{5}$$

Given the definitions of Z_L and Z_W that uses c_i, the steps in the prior proof are the same through equation 4, where the total losses among losers is equal to:

$$Z_L = |Z_L| = \frac{N}{2}|\bar{\epsilon}| + \frac{\Delta}{2}.$$

A Pareto improvement is impossible if losses among losers exceeds all of the welfare gains from efficiency enhancements (g_i) plus the gains from behavioral adjustments (b_i):

$$\frac{N}{2}|\bar{\epsilon}| + \frac{\Delta}{2} > \sum_i (g_i + b_i) \Leftrightarrow |\bar{\epsilon}| > 2(\bar{g} + \bar{b}) - \bar{\Delta}.$$

The statement in the proof uses the definition of $|\bar{\epsilon}|$. □

Condition 3. *Let $\tilde{c}_i$ be the initial costs from a policy, N be the number of agents, $\tilde{\mathbf{X}}_i$ be a vector of endogenous covariates observed by the planner consistent with a transfer scheme $T(\tilde{\mathbf{X}}_i)$, $T(\tilde{\mathbf{X}}_i)$ be a transfer scheme, $\bar{\Delta}$ the average funding gap, $\bar{\tau}^T$ be the average distortion from responses to the transfer function, and $\bar{g}$ be the average efficiency gain. If the average absolute targeting error exceeds twice the average efficiency gain minus the average funding gap; i.e.,*

$$\frac{1}{N} \sum_i |\tilde{c}_i - T(\tilde{\mathbf{X}}_i)| > 2\bar{g} - \bar{\Delta} - \bar{\tau}^T,$$

then there is no distribution of g with $g_i \geq 0$ $\forall i$ for which the policy and transfers create a Pareto improvement.

The steps in this proof are all exactly the same as for condition 1, except that the inefficiency loss from distortions, $\bar{\tau}^T$ acts as an increase in the revenue gap Δ. Replacing $\bar{\Delta}$ in the original proof with $\bar{\Delta} + \bar{\tau}^T$ is sufficient. □

B Appendix: Data comparisons

The CEX was chosen as the primary data source for this analysis because it includes a rich set of demographic covariates and a measure of gasoline expenditures, and because it is the standard data source in the most closely related literature. Gasoline expenditures, however, are based on self-reports and may be subject to mismeasurement. If there is a lot of noise in the expenditure data, this will make prediction more difficult. This section attempts to establish some sense of the reliability of CEX data by comparison to other surveys.

Of course, at the very highest level, problems of measurement do not challenge the key thesis of this paper. Instead, these problems reinforce it. If the best available data on expenditures are noisy measures of true burdens, it only makes it more difficult to design an accurate targeting scheme and thereby to compensate losers.

The National Household Travel Survey

An alternative measure of motor fuel consumption can be taken from the National Household Travel Survey, which is a nationally representative survey performed most recently in 2001, 2009 and 2017. That survey gathers a measure of annual vehicle miles traveled and then divides by the EPA estimated fuel economy of a vehicle to arrive at an estimate of annual fuel consumed. This is multiplied by average gasoline prices from the Energy Information Administration to impute expenditures. In contrast, the CEX asks consumers directly about expenditures.

The 2009 version of the NHTS is the most recent survey in which the miles traveled variable was based on two odometer readings (the survey respondent is asked to look at their odometer), rather than a retrospective self report. I compare the motor fuels expenditure data from that survey to the CEX from 2009. Figure 3 shows that fuel expenditure distribution from the two surveys for different samples. The top panel shows all households. This shows that the NHTS has higher expenditures on average, with a substantially longer right tail.

In part this may be due to differences in unit definitions across the two surveys, as the CEX is broken into smaller consumer units than the household definition used in the NHTS. Differences persist, however, when comparing households with the same number of members. The bottom two panels of Figure 3 compare households with one vehicle (on the left) and with two vehicles (on the right). In particular for the one vehicle households, the distributions do fit better. Nevertheless, the two data sources do show non-trivial differences in this fundamental measure.

Though there are some advantages to the measure of fuel expenditure in the NHTS, it has the disadvantages of requiring imputation of fuel economy and gasoline prices. Gasoline prices vary significantly across locations and time. Fuel economy varies substantially with where a vehicle is driven. Thus, it is not obvious which survey measure is more reliable. Regardless, the fact that there are substantial differences suggests that mismeasurement could be important.

Ultimately for the purposes of this paper, what matters is predictability. To compare predictability across the surveys I identify a set of demographic variables that appear to be defined consistently in both surveys: income, Census region, an urban indicator, family size

Figure 2: Comparison of Distribution of Implied Fuel Expenditure in CEX and NHTS

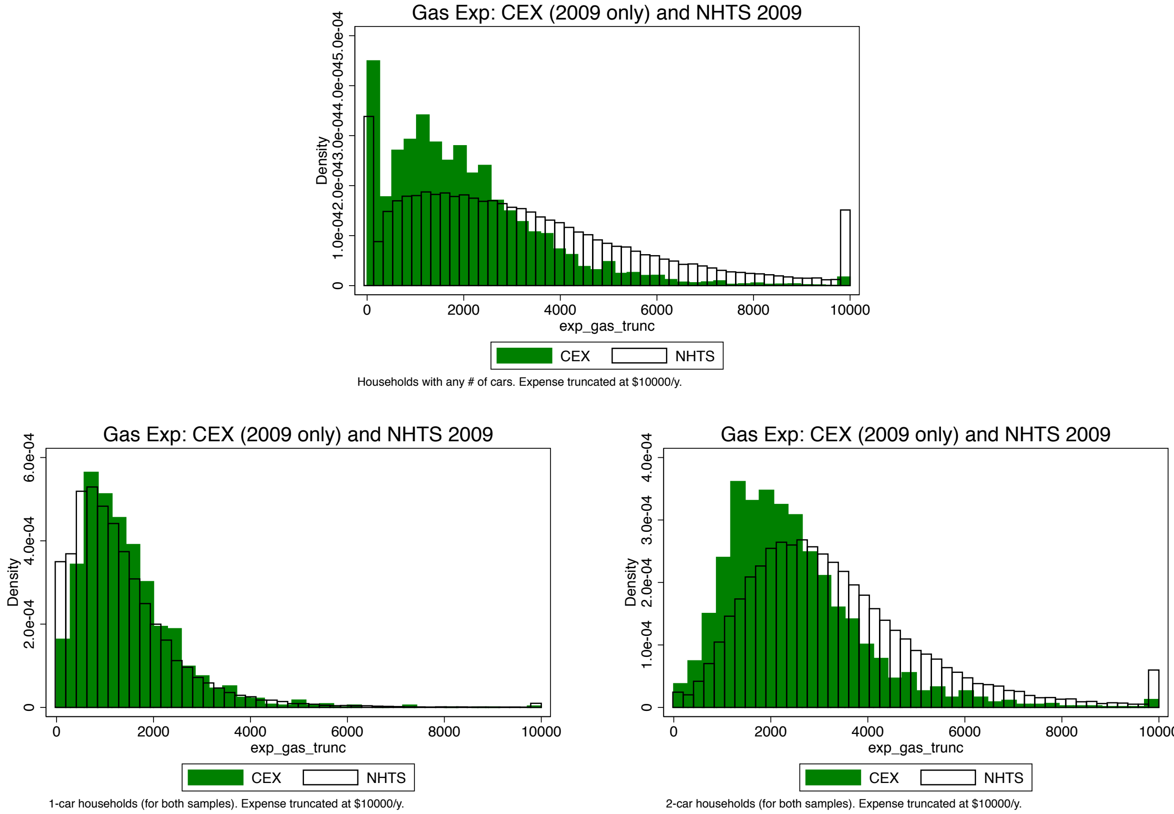


Figure shows histogram of estimated annual fuel expenditure by households using CEX and NHTS data. Left panel is for households with one vehicle. Right panel is for households with two vehicles. Both are from 2009 surveys. All distributions are truncated at $10,000 of annual expenditure.

and number of persons over 18. Table 9 reports the R^2 for parallel regressions of gasoline expenditures on these controls, varying the set of controls and whether the regressions using sample weights.

For the base set of controls, the CEX and NHTS show similar levels of predictability as summarized by the R^2. This is true regardless of weighting. In additional specifications (not shown), the results change very little when using dummies for the household size variables or adding state dummies instead of Census regions. The one difference that did emerge in a specification search was that the total number of vehicles owned by the household has a stronger explanatory power in the NHTS, and in particular when weighting, this variable notably increases prediction accuracy. The NHTS collects mileage information (from which expenditures are imputed) for each car, ensuring a mechanical connection. Table 9 reports the weighted and unweighted versions of these regressions showing the greater impact of vehicle controls in the NHTS.

Overall, the comparison of the CEX with the NHTS suggests that there are some notable differences in estimated gasoline expenditure, though in most cases there is not a large difference in predictability within the two samples. While mismeasured expenditures in the

Table 9: Predictability of Gasoline Expenditures in CEX versus NHTS

	CEX	NHTS	CEX	NHTS	CEX	NHTS	CEX	NHTS
R^2	.278	.232	.250	.267	.359	.380	.322	.431
Base controls	Y	Y	Y	Y	Y	Y	Y	Y
Number of vehicles	N	N	N	N	Y	Y	Y	Y
Weighted	N	N	Y	Y	N	N	Y	Y
N	9,116	137,938	9,116	137,938	9,116	137,938	9,116	137,938

Table compares 2009 CEX to 2009 NHTS. Dependent variable is annualized gasoline expenditures. Base controls include income, Census regions, urban dummy, family size and number of persons over 18. The additional variable is total number of vehicles in the household.

CEX may imply that the R^2 is artificially low as compared to some theoretical baseline, it is worth emphasizing a final time that trouble measuring consumption (and hence the burden of a tax) actually makes targeting transfers accurately more difficult.

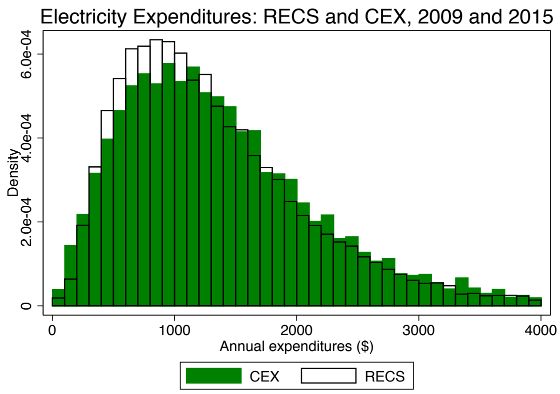
The Residential Energy Consumption Survey

This paper focuses on gasoline taxes, but it also briefly presents results on home energy consumption. The data quality of the home energy consumption variables in the CEX can be explored by comparison with the Residential Energy Consumption Survey (RECS), which is most recently available in 2009 and 2015. The RECS has the key advantage that electricity and natural gas expenditures are validated against billing records, so the data quality are much better for those variables than in most surveys.

Figure 3 shows the distribution of electricity and natural gas expenditures in the CEX and RECS, pooled for 2009 and 2015. Overall, the similarity in the distributions is broadly encouraging, but there are differences. The CEX shows more observations with low consumption, especially for gas. It also has a longer right tail. This may be in part because the CEX consumer units are on average smaller, but it may also be evidence of mismeasurement.

The primary concern with mismeasurement for the core purposes of this study is that it might artificially deflate the degree of predictability. Table 10 shows the R^2 from regressions with the overlapping common set of covariates between the RECS and CEX. The RECS does show a somewhat higher R^2. The data are not winsorized in these regressions. In other specifications (not shown), truncating the right tail of the distribution for high values has little effect on the R^2. Again, the high level point that consumption will be hard to measure and predict is reinforced if the CEX has measurement problems, though it certainly opens the possibility of using better measured surveys to design the transfer system.

Figure 3: Comparison of Distribution of Implied Electricity and Natural Gas Expenditures in CEX and RECS



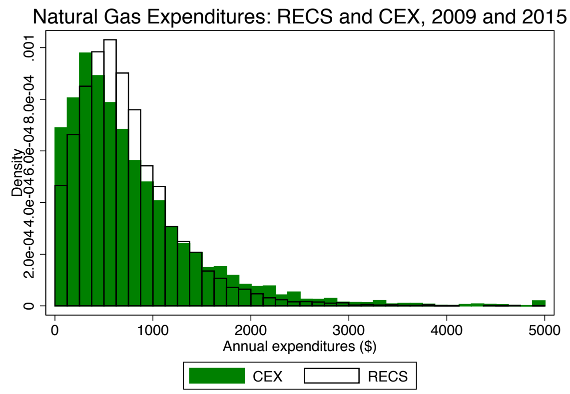


Figure shows histogram of estimated home energy expenditures by households using CEX and RECS data. Left panel is for electricity. Right panel is for natural gas. Both use 2009 and 2015 survey data combined.

Table 10: Predictability of Home Energy Expenditures in CEX versus NHTS

	Electricity				Natural Gas			
	CEX	RECS	CEX	RECS	CEX	RECS	CEX	RECS
R^2	.198	.262	.180	.263	.123	.218	.114	.178
Base controls	Y	Y	Y	Y	Y	Y	Y	Y
Weighted	N	N	Y	Y	N	N	Y	Y
N	17,802	17,769	17,802	17,769	11,263	10,798	11,263	10,798

Table compares 2009 and 2015 CEX to 2009 and 2015 RECS. Dependent variable is annualized expenditures on electricity or natural gas. Base controls include income, Census regions, urban dummy, family size and number of persons over 18. Samples are restricted to households with positive expenditures for natural gas.