Judging Nudging: Toward an Understanding of the Welfare Effects of Nudges Versus Taxes

John A. List, Matthias Rodemeier, Sutanuka Roy, and Gregory Sun

December 2022

Abstract

While behavioral non-price interventions ("nudges") have grown from academic curiosity to a bona fide policy tool, their relative economic efficiency remains under-researched. We develop a unified framework to estimate welfare effects of both nudges and taxes. We showcase our approach by creating a database of more than 300 carefully hand-coded point estimates of non-price and price interventions in the markets for cigarettes, influenza vaccinations, and household energy. While nudges are *effective* in changing behaviour in all three markets, they are not necessarily the most *efficient* policy. We find that nudges are more efficient in the market for cigarettes, while taxes are more efficient in the energy market. For influenza vaccinations, nudges may offer similar welfare gains to optimal vaccine subsidies. Importantly, two key factors govern the difference in results across markets: i) an elasticity-weighted standard deviation of the behavioral bias, and ii) the magnitude of the average externality. Nudges dominate taxes whenever i) exceeds ii). A policy mix that combines nudges and taxes dominates either alone.

JEL Codes: D61, D83, H21, Q41, Q48

Keywords: behavioral public policy, nudges, internality taxes, welfare, smoking, influenza vaccination, energy efficiency

---

Acknowledgments: David Franks went above and beyond with excellent research assistance and project management. Without his efforts the project would have never drawn to completion. We also thank Silvia Amalia Meneghesso for additional support.

Affiliations: List: Dept. of Economics, University of Chicago, NBER, and Australian National University. Rodemeier: Bocconi University. Roy: Australian National University. Sun: Olin Business School, Washington University. Rodemeier is the corresponding author at rodemeier.matthias@gmail.com.

1 Introduction

Non-price interventions motivated by insights from psychology, frequently referred to as "nudges," have received growing attention among academics and policymakers over the last decade. The extraordinary popularity of nudges has led to the implementation of many behavioral interventions across the globe attempting to shape individual behaviors. Nothing has seemingly been off limits, as today's choice architects commonly use social comparison nudges to help consumers conserve energy, warning labels on cigarette packages to deter smoking, and public campaigns to induce vaccinations. Many of these interventions are justified as cost-effective alternatives to traditional price and quantity regulations, such as taxes and subsidies or mandated quotas (Benartzi et al. 2017). The cost advantage over traditional policy tools is that the provision costs of nudges are often low, such that the change in behavior *per dollar spent on the intervention* is large.

While such a cost-based approach provides useful first insights into the comparison across policies, it does not take into account important factors of economic efficiency. Importantly, it does not quantify how the change in behavior caused by the nudge changes welfare to consumers and other members of society. A more complete benefit-cost analysis uses a revealed preference approach to understand how changes in behavior from various policy interventions map into welfare implications. There are few existing studies that include an applied welfare analysis of this type. By contrast, hundreds of empirical studies have estimated the reduced-form effects of nudges in many important markets of interest. We are therefore confronted with a vast amount of information on the efficacy of various nudges but have a far more limited understanding of their welfare impacts.

This challenge represents the motivation of our work. We develop an approach to estimate welfare effects of both nudges and taxes from the reduced-form treatment effects reported in hundreds of prior studies. In doing so, we provide the first comprehensive meta-analysis of welfare effects. Our methodology builds on the reduced-form approach to behavioral public policy evaluation due to Mullainathan, Schwartzstein, and Congdon (2012), and combines it with a large database of carefully hand-coded point estimates of non-price and price interventions. To provide a glimpse across key domains in which nudges are ubiquitous and behavioral biases are allegedly important,

we focus on data from three distinct markets: cigarettes, influenza vaccinations, and household energy consumption.

In total, we collect 311 point estimates on the effects of nudges and price interventions.[1] Our catalogue of studies covers a wide array of interventions including social norm comparisons, reminders and feedback, planning prompts and goal setting nudges, defaults, as well as any informational intervention. The intuition of our theory and nature of our empirical results can naturally be viewed within the three distinct markets we empirically explore.

In the market for cigarettes, we estimate that nudges, on average, increase the smoking cessation probability by 7.5% and reduce cigarette demand by 10%. We estimate an average price elasticity of -0.48, suggesting that the average nudge has the same effect on aggregate demand as a tax that increases the price of cigarettes by 20%. Leveraging our theoretical framework, we estimate the implied welfare impacts of nudges, of an optimal cigarette tax, as well as of a policy mix that combines both tools. We find that nudges cause a statistically significant increase in social welfare between \$0.35–\$1.12 per consumer per year. Importantly, nudges tend to outperform cigarette taxes for a wide range of auxiliary parameter values. The optimal cigarette tax amounts to \$6.50 per pack and raises welfare by \$0.36 per consumer. Interestingly, a policy mix that combines a nudge with a tax is only slightly superior in terms of welfare gains than the nudge in isolation.

The key intuition driving these results can be found in two important statistics: i) the elasticity-weighted standard deviation of the behavioral bias, and ii) the size of the average externality. While both nudges and taxes can correct the *average* behavioral distortion, they each have a unique comparative advantage in our framework. The comparative advantage of nudges lies in the reduction of heterogeneity in the bias, while the comparative advantage of taxes is the internalization of externalities. We show that nudges dominate taxes whenever i) is larger than ii). In the market for cigarettes, heterogeneity in the bias turns out to be a more important market distortion than externalities from smoking. This insight explains why nudges are more economically efficient than cigarette taxes.

[1] Since it is often difficult to precisely define what constitutes a nudge, we base our selection of nudge categories on Benartzi et al. (2017)

For the second market we consider, influenza vaccinations, we estimate that a nudge increases the vaccination take-up by, on average, 1.43%. The mean price elasticity is -0.33, which indicates that vaccine subsidies would have to decrease prices by 4.8% to generate the same effect as nudges. When estimating welfare effects, we find that the benefits of nudges over influenza subsidies are more limited than in the cigarette market. Nudges only dominate subsidies when price elasticities and nudge treatment effects are extremely negatively correlated. In the more likely scenario in which nudge and price effects are positively correlated, subsidies slightly outperform nudges. In this scenario, the nudge raises welfare by, on average, $74 per person, while the optimal vaccine subsidy increases welfare by $149 per person. The ratio between heterogeneity in bias to the average externality can again explain this result. The average positive externality of getting vaccinated is substantially larger than the standard deviation of the behavioral bias, which makes subsidies more efficient in this case. Yet, it is important to note that due to large standard errors, we cannot exclude the possibility that both policy tools yield equivalent welfare gains.

Finally, our most robust set of estimates come from the energy market. In terms of reduced-form effects, we find a mean average treatment effect of -5.4% on electricity consumption, and a mean price elasticity of -0.43. Thus, nudges have the same effect as a tax that raises the electricity price by 12.3%. We find that welfare gains from taxing electricity vastly exceed welfare gains from nudging: nudges increase social welfare by $49–$94 per household year, while the optimal tax raises welfare by $882 per household per year (the optimal tax equals $0.21 per kWh). For most parameter values, gains from taxation are 10-20 times larger than gains from nudging. Importantly, this result is independent of our specific assumption regarding how nudge and price effects are correlated. Furthermore, a policy mix that adds a nudge to the tax provides virtually no additional benefits over implementing a tax alone. The underlying mechanism for this stark result is that the negative externalities from electricity consumption (in the form of carbon emissions) are dramatically larger than the standard deviation of the behavioral bias. Thus, heterogeneity in biases is a relatively negligible source of friction in the electricity market when put into perspective.

Overall, our results show that, while nudges are *effective* in all three applications, they are not

always the most *efficient* intervention. The key statistics that predict when nudges dominate taxes are heterogeneity in the behavioral bias and the size of the average externality. As such, these two statistics can guide policymakers in choosing the most efficient policy instrument.

We view our combination of theory and empiricism as contributing to several unique strands of the literature. Our insights contribute to the nascent literature in behavioral public economics that studies optimal regulation in the presence of behavioral biases. We develop a theoretical model that builds on the frameworks in Bernheim and Rangel (2009), Mullainathan, Schwartzstein, and Congdon (2012) and Farhi and Gabaix (2020) and provide useful extensions. Our empirical results add to a small set of papers that quantify the welfare effects of behavioral public policies. Prior studies have estimated the welfare effects of nudges and similar non-price interventions (Chetty, Looney, and Kroft 2009, DellaVigna, List, and Malmendier 2012, DellaVigna et al. 2016, Allcott and Kessler 2019,Rodemeier 2020, Butera et al. 2022, Löschel, Rodemeier, and Werthschulte 2022, Rodemeier and Löschel 2022), as well as of behaviorally-motivated taxes (Allcott, Mullainathan, and Taubinsky 2014, Allcott and Taubinsky 2015, Allcott, Lockwood, and Taubinsky 2019). Our paper makes a unique contribution to this literature by offering the first meta-analysis of welfare effects of both nudges and taxes, while relying only on reduced-form effects from prior studies.

Our structural estimates of behavioral parameters also contribute to the emerging literature in "structural behavioral economics" that identifies structural parameters proposed in theories at the intersection of psychology and economics (DellaVigna 2018). This literature has provided insights across a myriad of topics, including estimating discount functions over the life-cycle (Laibson, Repetto, and Tobacman 2007), measuring the nature of risk preferences (see, e.g., Barseghyan et al. 2013) and projection bias (Conlin, O'Donoghue, and Vogelsang 2007), as well as exploring gift exchange at work (DellaVigna et al. 2022) and why firms engage in corporate social responsibility (Hedblom, Hickman, and List 2019).

From an applied policy perspective, our welfare results speak to an extensive literature on the role of behavioral interventions in health and energy policy. An interdisciplinary field involving medicine, psychology, and economics has studied how nudges can help people to stop smoking

(e.g., Armitage and Arden 2008, Hennrikus et al. 2005, Borland, Balmford, and Swift 2015), or get vaccinated (e.g., Milkman et al. 2011, Srinivasan et al. 2020, Frank, McMurray, and Henderson 1985). Another large number of studies has investigated how nudges can be used to reduce households' energy consumption (e.g., Allcott 2011, Jessoe and Rapson 2014, Allcott and Wozny 2014, Allcott and Rogers 2014, Houde 2018, Andor et al. 2018, Allcott and Knittel 2019, Löschel, Rodemeier, and Werthschulte 2022). Our paper complements these studies by offering an insight into the welfare effects of these interventions.

Finally, our paper offers a novel database of reduced-form effects of behavioral interventions and relates to a growing literature using meta analyses on nudging (Benartzi et al. 2017, Antinyan and Asatryan 2019, Hummel and Maedche 2019, DellaVigna and Linos 2022). Different from prior contributions, however, we leverage reduced-form estimates to draw insights on efficiency effects of various policy approaches and highlight how methodologies in this spirit can provide deep implications for the choice of the optimal policy tool.

The remainder of our study is structured as follows. In Section 2, we show theoretically how to evaluate the welfare effects of nudges and taxes based on reduced-form treatment effects. Section 3 describes the empirical estimation of the welfare formulae. We discuss the data collection in Section 4. In Section 5, we discuss both reduced-form estimates and the implied welfare effects. Section 6 concludes.

2 Theoretical Framework

In this section, we introduce a simple model that quantifies the welfare effects of nudges and taxes. We build on the frameworks in Bernheim and Rangel (2009), Mullainathan, Schwartzstein, and Congdon (2012) and Farhi and Gabaix (2020) and revise important results. While we also add some new insights, our main contribution is empirical. We briefly discuss the relevant welfare formulae and show how they motivate the empirical analysis. For a more in-depth discussion of the theoretical results, we refer the reader to the literature above.

We assume throughout that the (rational benchmark) demand side of the model is determined by a private benefit function $V(q)$. This framework includes any model where demand is derived by adding up a population of agents with quasi-linear utility, in which case

$$V(q) = \max_{q_1,\ldots,q_N} \sum_{i=1}^{N} u(q_i) \quad \text{s.t.} \quad \sum_{i=1}^{N} q_i = q$$

The supply side of the model is characterized by a constant returns production function with marginal cost c. Firms are assumed competitive, so given a tax t, consumers will face price $p = c+t$. Given our constant returns assumption, the supply side of the market is perfectly elastic, so firms in equilibrium earn zero profits.

Finally, we add two market frictions to the model that motivate policy making: externalities and internalities. The marginal externality is denoted by ξ and assumed constant. The internality, also referred to as the behavioral bias, is given by b_n, and is a function of a binary nudge $n \in \{0, 1\}$. A value of $b_n \neq 0$ affects the *behavioral* aspects of the model by making consumers systematically misperceive the marginal benefit of a unit of consumption. Formally, demand is characterized by $V'(q)+b_n = p$, rather than the usual first-order condition $V'(q) = p$. Consumers overvalue benefits of consumption whenever $b_n > 0$, and undervalue them whenever $b_n < 0$. This is a common feature in behavioral welfare economics: Biases do not enter experienced utility but affect choices, thereby creating a wedge between marginal utility and price. Conversely, the externality ξ does not affect choices but directly enters the social welfare function.

We allow for heterogeneity by letting the private benefit V be a random function and letting internalities and externalities, b and ξ, be random variables, which also may co-vary which each other. We assume that the government retains a neutral budget and returns taxes to consumers in lump sum.

For any given realization of market parameters, social welfare is then given by

$$W(q) = \underbrace{V(p) - (c+t)q(t,n)}_{\text{Consumer Surplus}} + \underbrace{tq(t,n)}_{\text{Governemnt Revenue}} - \underbrace{\xi q(t,n)}_{\text{Externality}} \tag{1}$$

$$= V(p) - cq(t,n) - \xi q(t,n) \tag{2}$$

The (infeasible) first-best allocation of an omniscient social planner satisfies the condition

$$W'(q) = 0 \iff V'(q) = c + \xi \tag{3}$$

While we do not view this as a realistically atainable benchmark, it is a useful technical reference point for some of the below calculations. More relevant to the real world, when tax is t and the nudge is n, equilibrium quantities solve the first order condition

$$V'(q) = c + t - b_n. \tag{4}$$

Thus, consumption deviates from the first-best allocation because 1) consumers ignore social costs, and 2) consumers misperceive private benefits. Unless these two market frictions coincidentally cancel each other out ($\xi = -b_n$ with probability 1), there is room for welfare enhancing policy interventions. We next consider how taxes and nudges can alleviate market distortions and how we can measure their welfare effects empirically.

2.1 Quantifying the Welfare Effects of Nudges and Taxes

Throughout this section and the rest of the paper, we make one final and common assumption, which can alternatively be thought of as an approximation. We assume that demand functions, $D(t,n)$, are linear. Linearity amounts to the assumption that $V''(q) = V''$ is constant since $D'(t,n) = (V'')^{-1}(t,n)$. By the assumption that $D'(t,n) = D'$ is constant, V is quadratic, and so is W. A more general interpretation of our results is that the following derivations can be viewed as

second-order Taylor approximations of welfare effects under any arbitrary welfare function that can be represented by equation (2). Taking the expectation of the individual demand functions gives aggregate demand, $\mathbb{E}[D(t,n)]$.

Let W^* be expected welfare given the first-best allocation described in the previous section, and let q^* solve $V'(q^*) = c + \xi$, so that $W^* = \mathbb{E}[V(q^*)]$. Since the welfare function is quadratic, we must have that $W(q) = W^* + \frac{1}{2}\mathbb{E}[V''(q-q^*)^2]$. Additionally, our linear demand assumption implies that q is a linear function of t and n and is given given by $q = q^* - \frac{t-(b_n+\xi)}{V''}$, and that $D' = (V'')^{-1}$. We can thus alternatively write welfare as a function of taxes and nudges,

$$W(t,n) = W^* + \frac{1}{2}\mathbb{E}[(t-(b_n+\xi))^2 D']. \tag{5}$$

The realized welfare loss from market frictions depends quadratically on the wedge between price and social cost and has magnitude increasing in the slope of the demand curve. This result is intuitive: Where the demand curve has a larger slope, the social welfare function is more curved, so the welfare loss from missing society's first order condition is higher per unit error. An alternative way of putting this is to first note that a given realization of $b+\xi$ causes a larger deviation from the optimal quantity if consumers are more price elastic. Higher price elasticities, for a given value of $b+\xi$, therefore, imply a larger deadweight loss.

The weighting interpretation is sufficiently important that we introduce some notation to accommodate it. Define a set of weights as $W \equiv \frac{D'}{\mathbb{E}[D']}$. Define the weighted mean $\mathbb{E}_W$ by $\mathbb{E}_W[X] = \mathbb{E}[WX]$ for some random variable X. Similarly, define the weighted variance by $\mathrm{Var}_W(X) = \mathbb{E}_W[X^2] - \mathbb{E}_W[X]^2$. We can now express equation (5) as

$$W(t,n) = W^* + \frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[(t-(b_n+\xi))^2]. \tag{6}$$

We refer to welfare with no tax and no nudge, $W(0,0)$, as our *baseline*, and then analyze welfare effects of nudges and taxes relative to this baseline. $W(0,0)$ is given by

$$W(0,0) = W^* + \frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[(b_0 + \xi)^2]. \tag{7}$$

2.1.1 Welfare Effects of a Nudge

The effect of a nudge depends on by how much it reduces the internality. Throughout this paper, we assume that nudges remove as proportion θ of the bias, $b_1 = \theta b_0$. Welfare under a nudge is

$$W(0,1) = W^* + \frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[((1-\theta)b_0 + \xi)^2]. \tag{8}$$

Note that if the nudge is fully de-biasing, $\theta = 1$, the remaining deadweight loss relative to first best is a function of the mean and variance of the externality, $W(0,1) = W^* + \frac{1}{2}\mathbb{E}_W[\xi^2]\mathbb{E}[D'] = W^* + \frac{1}{2}\mathbb{E}[D'](\mathbb{E}_W[\xi]^2 + \mathrm{Var}_W[\xi])$.

The welfare effect of a nudge relative to baseline is given by

$$\Delta_n W(0,1) = -\frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[\theta(2-\theta)b_0^2 + 2\theta b_0 \xi]. \tag{9}$$

In the next sections, we will compare the welfare effect of a nudge given here to the welfare effects of alterantive sections.

To gain some intuition for Equation (9), it is helpful to consider the case where $\theta = 1$. In this case, the welfare effect becomes

$$\Delta_n W(0,1) = -\frac{1}{2}\mathbb{E}[D']\left\{\mathbb{E}_W[b_0(b_0 + 2\mathbb{E}_W[\xi])] + 2\mathrm{Cov}_W(b_0, \xi)\right\} \tag{10}$$

The first term $\mathbb{E}_W[b_0(b_0 + 2\mathbb{E}_W[\xi])]$ is the welfare change due to correcting the internality in the presence of a deterministic pre-existing distorition induced by the externality of average magnitude $\mathbb{E}_W[\xi]$. The form of this expression is remniscent of the classical result of (Harberger 1964) that the effect of a marginal change in tax on welfare is equal to the effect of the tax *times* the pre-existing level of distortion. Thus, the level of externality in the market determines the quantitative

importance of the nudge in improving welfare.

To gain intuition for how the presence of externalities affects the welfare implications of the nudge, we assume for the moment that $b_0 > 0$ (the opposite case, $b_0 < 0$ is symmetric). Consider first the case where $\mathbb{E}_W[\xi] > 0$ as well. In that case, the bias and externality push in the same direction, causing more extreme deviations from the social first order condition relative to the case of only one of the two distortions alone. Since welfare losses are quadratic in the size of the distortion, correcting a bias that amplifies a pre-existing externality has more positive welfare impacts relative to correcting an equally sized bias in an otherwise undistorted market. Alternatively, if $\mathbb{E}_W[\xi] < 0$, then the bias and externality push in opposite directions, so the bias partially offsets the existing externality. In this case, the benefits of correcting the bias are partially offset by the costs of no longer cancelling out the externality, so welfare effects will be less positive than correcting an equal sized bias in an otherwise undistorted market (and can even lead to a net negative effect).

The second term $2\mathrm{Cov}(b_0, \xi)$ reflects further adjustments to welfare due to heterogeneity in how bias and externality interact. If biases tend to be large when externalities are also large (and in the same direction as the bias), then the marginal benefit of reducing bias is especially high, again, due to the fact that the welfare cost of distortions is quadratic in the distortion.

2.1.2 Welfare Effects of a Tax

Suppose that the government instead decides to use a tax but no nudge. To optimize the choice of tax, we differentiate the expected welfare criterion with respect to t, which yields:

$$\mathbb{E}\left[(t - (b_n + \xi))D'\right] = 0, \tag{11}$$

which implies that optimal t must satisfy

$$t_0^* = \mathbb{E}_W[(b + \xi)]. \tag{12}$$

The optimal tax equals a weighted average of the sum of the market frictions, where the weights

are proportional to the slope of the demand curve. The optimal tax puts larger weights on more price elastic consumer groups, because larger demand elasticities imply larger deviations from the optimum for any given value of $b+\xi$. This result, again, highlights the importance of the weighting and has previously been established by Allcott and Taubinsky (2015).

Consider now, the welfare effect of the optimal tax. Substituting the optimal tax into $W(t,0)$ shows that welfare under this policy is $W(t_0^*,0) = W^* + \frac{1}{2}\text{Var}_W(b_0+\xi)\mathbb{E}[D']$. This illustrates that once the government has corrected the expected sum of the market frictions, the remaining deadweight loss stems from heterogeneity which is determined by the weighted variance of market frictions, $\text{Var}_W(b_0+\xi)$ and the aggregate demand elasticity. Since the tax cannot be targeted—i.e. cannot be set equal to each realization of $b_0+\xi$—it cannot fully correct the choice of each consumer, and therefore cannot achieve the first-best allocation. As we discuss later, this is where the nudge may have an important comparative advantage over the tax.

The welfare effect from the optimal tax relative to baseline with no tax and no nudge is $\Delta_t W(0,0) = W(t_0^*,0) - W(0,0)$ and given by

$$\Delta_t W(0,0) = -\frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[b+\xi]^2 \tag{13}$$

$$= -\frac{1}{2}\frac{\left(\mathbb{E}[D']\mathbb{E}[b+\xi] + \text{Cov}(D', b+\xi)\right)^2}{\mathbb{E}[D']} \tag{14}$$

The welfare effect of the optimal tax increases quadratically in the weighted exptectation of the market frictions. This expectation can be decomposed into the unweighted expected sum of the market frictions, and the covariance between these frictions with demand elasticties. If $\text{Cov}(D', b+\xi) = 0$, the welfare effect simplifies to $\Delta W_t(0,0) = -\frac{1}{2}\mathbb{E}[b+\xi]^2\mathbb{E}[D']$.

Before deriving the other welfare effects, it is also worthwhile to point out the relationship between the above formulae and the prior sufficient statistics literature. Specifically, consider the case where there are no externalities ($\xi = 0$) and where nudges are fully effective, ($\theta = 1$). In this case, the optimal tax is given by $t_0^* = \mathbb{E}_W[b_0]$, while the welfare impact of this policy is $-\frac{1}{2}\mathbb{E}[D']\mathbb{E}_W[b_0]^2$.

Thus, $\mathbb{E}_W[b_0]$ is a key quantity of interest, and we aim to understand it better.

Some simple algebraic manipulations show that $\mathbb{E}_W[b_0] = \frac{\mathbb{E}[D'b_0]}{\mathbb{E}[D']}$. Given our assumption of how nudges work, the consequence of a fully effective nudge is to change perceived prices by amount b_0, which therefore changes quantity demand by exactly the quantity $D'b_0$. Thus, letting T be the *treatment effect* of a nudge, we see that in fact, $\mathbb{E}_W[b_0] = \frac{\mathbb{E}[T]}{\mathbb{E}[D']}$. The RHS of this final expression is a well known quantity in the literature and is commonly referred to as the *equivalent price metric* (EPM). Our framework gives an interpretation of the EPM as a particular sufficient statistic. Specifically, under fairly weak restrictions on heterogeneity (i.e., linear demand functions), it corresponds to the optimal uniform tax when $i)$ markets are perfectly competitive, $ii)$ nudge policies are unavailable, or infeasible at scale, and $iii)$ there is no cost of raising government funds. While these assumptions are restrictive in general, the discussion nonetheless clarifies that the EPM is not an arbitrary statistic but instead corresponds to a (very specific) policy relevant parameter.[2]

2.1.3 Welfare Effects of Taxes and Nudges in Combination

It is interesting to compare the welfare effects of a the optimal tax to the welfare effects of a nudge. To isolate the key tradeoffs, it is helpful to consider the best-case scenario for nudges where they are perfectly effective at debiasing ($\theta = 1$). In this case, using the formulas derived in this and the previous subsection, we have that the difference between welfare under an optimal tax compared to welfare under a nudge is given by

$$W(t_0^*, 0) - \Delta_n W(0, 1) = \frac{1}{2}\mathbb{E}[D'] \left(\mathrm{Var}_W(b_0) + 2\mathrm{Cov}_W(b_0, \xi) - \mathbb{E}_W[\xi]^2\right)$$

This shows that a nudge is superior to a tax if and only if

$$\mathrm{Var}_W(b_0) + 2\mathrm{Cov}_W(b_0, \xi) \geq \mathbb{E}_W[\xi]^2$$

[2] Allcott and Taubinsky (2015) show that EPM statistic may deviate from $\mathbb{E}_W[b]$ when the density of marginal consumers with some realization of b changes along the aggregate demand curve. Our assumption of linear demand for each realization of b rules out this possibility. While this might be restrictive, linearity is a standard assumption in the literature on sufficient statistics and structural estimation.

This formula is somewhat easier to interpret when internalities and externalities are independent of one another, in which case, the above inequality reduces to the condition that

$$\mathrm{sd}_W(b_0) \geq |\mathbb{E}_W[\xi]|, \tag{15}$$

where sd_W is the W-weighted standard deviation. In other words, nudges are superior to taxes if and only if the degree of heterogeneity in bias is larger than average externality.

The above discussion highlights that nudges and taxes fundamentally target two separate sources of market inefficiency. Under our framework, nudges act implicitly like a first-degree price discrimination tool, surgically targeting consumers and correcting their individual-specific biases. It therefore tends to work best when biases are heterogeneous, so that the value of this implicit first-degree price discrimination is high. On the other hand, within our framework, biases have no ability to affect the externality component of an individual's decision calculus, so it is unable to affect any aspects of market friction that are mediated through externalities. By contrast, taxes have complementary strengths and weaknesses. Because taxes operate directly on prices, they can address both internally and externality aspects of market frictions. However, they are a coarse policy instrument which are only able to correct the *average* level of market frictions. It therefore tends to have a comparative advantage when the average externality is high, as this is precisely the component of market frictions that nudges are not able to address.

In light of our discussion, a policy that uses taxes and nudges in conjunction can obtain the best of both worlds, as they are able to compensate for each other's shortcomings. We therefore consider a policy mix in which the government uses both the tax and the nudge. In particular, we can improve upon a nudge with a tax that corrects the remaining market distortions after the nudge has been implemented. This is operationalized by the tax $t_1^* = \mathbb{E}_W[(1-\theta)b_0+\xi]$. Welfare under this policy is given by $W(t_1^*, 1) = \frac{1}{2}\mathbb{E}[D']\mathrm{Var}_\mathrm{W}((1-\theta)b_0+\xi)$, implying that the remaining deadweight loss relative to first best again stems from heterogeneity in the (post-nudge) bias and the externality.

The welfare gain from this combination of taxes and nudges relative to status quo is

$$\Delta_{tn}W(0,0) = -\frac{1}{2}\mathbb{E}[D'](\mathbb{E}_W[\theta(2-\theta)b_0^2 + 2\theta b_0\xi] + \mathbb{E}_W[(1-\theta)b_0 + \xi]^2). \tag{16}$$

Again, it is helpful to consider the case where the nudge is fully debiasing, $\theta = 1$. The welfare effect can then be simplified as

$$\Delta_{tn}W(0,0) = -\frac{1}{2}\mathbb{E}[D']\{\mathbb{E}_W[(b_0 + \mathbb{E}_W[\xi])^2]\mathrm{Cov}(b_0,\xi)\}. \tag{17}$$

As before, the RHS splits into economically interpretable terms. The first term, $\mathbb{E}_W[(b_0+\mathbb{E}_W[\xi])^2]$, is the expected welfare gain from fixing the net effect of a deterministic externality $\mathbb{E}_W[\xi]$ plus a stochastic interality, b_0. The second term, $\mathrm{Cov}_W(b_0,\xi)$, already shows up in the expression for the welfare effect of a isolated nudge.

3 Empirical Implementation

In this section we show how to connect the theoretical framework from the previous section to empirical observations. Specifically, in Subsection 3.1, we describe the form of our meta-analysis datasets and discuss the key identification and estimation challenges in linking our theoretical framework to the available data. In Subsection 3.2, we describe our concrete algorithm for translating the discussion in Subsection 3.1 into estimated welfare impacts of nudges and taxes.

3.1 Identification and Estimation

3.1.1 Meta-analysis data

The discussion in this section should be understood as describing our empirical strategy for a single market (e.g., cigarettes). We implement the formulae derived in this section three times, one for each market we study. For each fixed market, we have access to two meta-analysis datasets, which we will refer to as N and Π, corresponding respectively to *nudge* and *price* treatments. Each study $n \in N$ estimates the effect of some nudge on quantities while each study $\pi \in \Pi$ estimates the

effect of price on quantity. For each nudge study $n \in N$, we record an estimated percent treatment effect, $\hat{T}_n$ as well as a reported standard error, $\hat{\sigma}_n$. On the other hand, each pricing study $\pi \in \Pi$ corresponds to the effect of a price change on quantities. For these studies, we record an estimated demand elasticity, $\hat{\varepsilon}_\pi$ as well as a reported standard error, $\hat{\sigma}_\pi$.

We assume that our meta-analysis datasets are generated as follows. Each study $n \in N$ corresponds to a realization of the market, so each study is associated with a draw from the joint distribution of demand curves and nudges: $(D_n, b_n) \sim F_{D,b}$. Given these realizations of market parameters, the "true" percent treatment effect for the study is given by $T_n = \theta \frac{D'_n}{D_n} b_n$. The estimated treatment effects and standard errors are then drawn in such a way that $\hat{T}_n \sim \mathcal{N}(T_n, \hat{\sigma}_n)$.

We assume that data from the price study are generated in a similar manner. Each study $\pi \in \Pi$ corresponds to a realization of the market and is thus associated with a draw $(D_\pi, b_\pi) \sim F_{D,b}$. The true elasticity is thus given by $\varepsilon_\pi = \frac{pD'_\pi}{D_\pi}$, while estimated treatment effects and standard errors are drawn so that $\hat{\varepsilon}_\pi \sim \mathcal{N}(\varepsilon_\pi, \hat{\sigma}_\pi)$.

3.1.2 Identification of Nudge Treatment Effect and Elasticity Marginal Distributions

Given the data-generating process for N and Π described above, standard deconvolution techniques can be used to show that the marginal distributions F_T and F_ε of treatment effects T and elasticities ε are non-parametrically identified, assuming a large dataset of studies.[3] In practice, however, we have found the non-parametric approach to be too demanding on data relative to our sample sizes. As a result, we have opted for a more parametric approach. Specifically, we assume that $T_n \sim \log \mathcal{N}(\mu_T, \sigma_T^2)$ while $\varepsilon_n \sim \log \mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$. In this case, the law of total expectation shows that $\mathbb{E}[\hat{T}_n] = \mathbb{E}[T_n]$ while the law of total variance implies that $\mathrm{Var}(\hat{T}_n) = \mathrm{Var}[T_n] + \mathbb{E}[\mathrm{Var}_n(T_n)] = \mathrm{Var}[T_n] + \mathbb{E}[\hat{\sigma}_n^2]$. Thus, the sample mean of estimated treatment effects is a consistent estimate of the average treatment effect while the sample analogue of the ANOVA identity consistently estimates the variance of treatment effects in the population. The same analysis shows how the mean and variance of the distribution of elasticities can be consistently estimated. Because there is an

[3] See, for instance, Stefanski and Carroll (1990)

injective mapping from means and variances into the parameters, μ, σ^2 of a log-normal distribution, the method of moments allows us to translate these estimates of $\mathbb{E}[T_n], \mathbb{E}[\sigma_n^2], \mathbb{E}[\varepsilon_\pi], \mathrm{Var}[\varepsilon_\pi^2]$ into estimates of the parameters of the log-normal distribution, $\mu_T, \sigma_T^2, \mu_\varepsilon, \sigma_\varepsilon^2$.

3.1.3 Non-identification of Welfare Formulae

Above we saw that the marginal distribution of treatment effects F_T and the marginal distribution of elasticities, F_ε are identified using our meta-analysis data. However, looking at welfare formulae (9), (14), and (16), it is clear that the joint distribution of (b_0, ε, ξ), along with the nudge effectiveness parameter θ are all needed to identify the welfare effects of the various policies under consideration. This implies that our meta-analysis data alone does not suffice to point identify the effects we aim to study without additional structure. In this section, we therefore list out what information is missing and how we can use auxilliary information, structural assumptions, and sensitivity analyses to overcome the resulting non-identification issues.

First, since we are focusing on performing a meta-analysis of nudge effects, we do not directly analyze data about the marginal distribution of externalities ξ. Instead, we rely on relevant prior literatures to provide us with estimates of the distribution of ξ.

Second, even given knowledge about the marginal distribution of ξ, we do not know how ξ jointly varies with b_0 and ε. In the contexts of the particular markets we study, externalities typically reflect broader macro-economic impacts on participants even outside of the markets we study. By contrast, ε and b_0 are parameters reflecting individual-level demand for participants specifically participating in the markets we are analyzing. Since the underlying processes generating ξ and (b_0, ε) come from fundamentally different sources, we argue that to first approximation, it is reasonable to treat externalities as exogenous and assume that $\xi \perp (b_0, \varepsilon)$.

Third, the studies in our meta-analysis seldom include price and nudge treatments within the same experiment. As a result, we are not able to identify the dependence structure between b_0 and ε. We therefore perform a sensitivity analysis to test how robust our results are to our assumptions about how b_0 and ε are related. Specifically, in the next subsection, we will describe how we pa-

rameterize the degree of correlation between nudge treatment effects and price elasticities, which in turn parameterizes the degree of correlation between internalities and price elasticities.

Finally, the nudge effectiveness θ is difficult to measure and not addressed in most studies. The standard approach in structural behavioral economics is to assume that the nudge is fully effective ($\theta = 1$), such that its treatment effect can be used to identify the magnitude of the behavioral bias. As with the dependence between b_0 and ε, in our empirical exercises, we will compute our welfare estimates under various assumptions about θ and show that our key conclusions do not rely on the particular assumption we make about the value of θ.

3.2 Welfare Formula Implementation Details

In the previous subsection, we modeled the marginal distributions of T and ε as log-normal and estimated their parameters. In this section, following the identification discussion sketched above, we model the joint distribution of T, ε, ξ as

$$\begin{pmatrix} T \\ |\varepsilon| \\ \xi \end{pmatrix} \sim \log\mathcal{N}\left(\begin{pmatrix} \mu_T \\ \mu_\varepsilon \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} \sigma_T^2 & \rho\sigma_T\sigma_\varepsilon & 0 \\ \rho\sigma_T\sigma_\varepsilon & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix}\right)$$

where $\rho = [-1, 1]$ is a sensitivity parameter dictating the correlation between the treatment effect and elasticity. Let p be the price of the good and let q be the baseline quantity demanded. Then assuming some value of θ, bias, B is given by $B = \frac{pT}{\theta\varepsilon}$ while slope, S is given by $S = \varepsilon\frac{q}{p}$. Applying the definition of the log-normal distribution shows that the joint distribution of B, S, ξ is again a log-normal distribution given by

$$\begin{pmatrix} B \\ |S| \\ \xi \end{pmatrix} \sim \log\mathcal{N}\left(\begin{pmatrix} \mu_B \\ \mu_S \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} \sigma_B^2 & \sigma_{BS} & 0 \\ \sigma_{BS} & \sigma_S^2 & 0 \\ 0 & 0 & \sigma_\xi^2 \end{pmatrix}\right)$$

where $\mu_B = \mu_T - \mu_\varepsilon + \log\frac{p}{\theta}$, $\mu_S = \mu_\varepsilon + \log qp$, $\sigma_B^2 = \sigma_T^2 - 2\rho\sigma_T\sigma_\varepsilon + \sigma_\varepsilon^2$, $\sigma_S^2 = \sigma_\varepsilon^2$, and $\sigma_{BS} = \rho\sigma_T\sigma_\varepsilon - \sigma_\varepsilon^2$.

In Appendix A, we derive the following closed form expressions for various welfare formulae in terms of the parameters of the log-normal distribution, $\mu_T, \mu_\varepsilon, \mu_\xi, \sigma_T^2, \sigma_\varepsilon^2, \sigma_\xi^2, \rho$ as well as the nudge-effectiveness parameter θ:

$$\begin{aligned}
W^* - W(0,0) &= \frac{1}{2}\{\exp\left(2\mu_\xi + \mu_S + [4\sigma_\xi^2 + \sigma_S^2]/2\right) \\
&\quad + \exp(\mu_B + \mu_S + \mu_\xi + [\sigma_B^2 + 2\sigma_{BS} + \sigma_S^2 + \sigma_\xi^2]/2) \\
&\quad + \exp\left(2\mu_B + \mu_S + [4\sigma_B^2 + 4\sigma_{BS} + \sigma_S^2]/2\right)\} \\
W^* - W(t_0^*,0) &= \frac{1}{2}\{[\exp(\sigma_\xi^2) - 1]\exp(2\mu_\xi + \mu_S + \sigma_\xi^2 + \sigma_S^2/2) \\
&\quad + \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\
&\quad - \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 2\sigma_B^2]/2)\} \\
W^* - W(0,1) &= \frac{1}{2}\{(1-\theta)^2\exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\
&\quad + 2(1-\theta)\exp(\mu_S + \mu_B + \mu_\xi + [\sigma_S^2 + 2\sigma_{BS} + \sigma_B^2 + \sigma_\xi^2]/2) \\
&\quad + \exp(\mu_S + 2\mu_\xi + [\sigma_S^2 + 4\sigma_\xi^2]/2)\} \\
W^*(t_1^*,1) &= \frac{1}{2}\{[\exp(\sigma_\xi^2) - 1]\exp(2\mu_\xi + \mu_S + \sigma_\xi^2 + \sigma_S^2/2) \\
&\quad + (1-\theta)^2[\exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\
&\quad - \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 2\sigma_B^2]/2)]\}
\end{aligned} \tag{18}$$

We have therefore derived closed form expressions for the various welfare formulae in terms of the parameters of the log-normal distribution, $\mu_T, \mu_\varepsilon, \mu_\xi, \sigma_T^2, \sigma_\varepsilon^2, \sigma_\xi^2, \rho$ as well as the nudge-effectiveness parameter θ. Recall that while the parameters $\mu_T, \mu_\varepsilon, \sigma_T^2, \sigma_\varepsilon^2$ come from our meta-analysis point estimates, and the parameters μ_ξ, σ_ξ^2 come from other literature concerning externality estimates, the final two parameters ρ, θ are unknown and hence will be subjected to a sensitivity analysis. For each value of ρ, θ under consideration, the entire vector of parameters are plugged into the above welfare formulae to get a single estimate of the effects of the various policies under consideration.

4 Inclusion Criteria and Data Collection

In collecting data we followed the best practices outlined in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al. 2022).

4.1 Inclusion Criteria

There were a few implicit criteria for a paper to be included in our meta-analysis. First, as English was the common language of the research team collecting the data, only papers in English language were included. Second, papers had to be digitally accessible. We are, therefore, more likely to sample newer papers in English-speaking countries.

We chose three important markets in which behavioral public policies are ubiquitous and for which a high number of nudge studies was available: the markets for cigarettes, vaccines, and household electricity. We focus on consumer markets for which our theoretical framework can be easily applied. We refrain from studying topics related to human capital formation, labor supply and savings behavior, as these applications cannot be captured by the our simple model without additional extensions.

We straightforwardly define a "price intervention" as any policy that changes the sales price of the underlying good (i.e., of cigarettes, vaccines, and electricity). On the other hand, the definition of a "nudge" in the broader literature is more fuzzy. One might argue that any non-price intervention could be defined as a nudge. However, strictly speaking this would also imply that medication or therapy sessions with a psychologist (e.g., to reduce nicotine addiction) would be included in our analysis, which seems intuitively inapplicable to the existing literature on nudges. To avoid arbitrariness in the selection process, we follow the paper by Benartzi et al. (2017) in the choice of interventions we classify as a nudge. This includes many informational interventions, such as social norms nudges that inform consumers about the consumption of their peers. Another example are usage information about the consequences of consumption on health, private costs, or costs on society as a whole. Similarly, reminders and feedback nudges are included, such as appointment

reminders, real-time feedback, or billing information. The selection also includes all forms of interventions that ask consumers to make plans for future consumption, such as goal setting prompts and pre-commitment nudges. Finally, the selection includes "streamlining nudges" that simplify tasks, as well as "default nudges" that change the pre-selection of a choice set.

We deviate from Benartzi et al. (2017) in that they do not count "educational interventions" as nudges. The reason we include such interventions is that there is no clear distinction between what constitutes information provision (a form of nudge in Benartzi et al. 2017) and an educational intervention. For instance, a leaflet that informs about the health risks of smoking may be defined as both an informational nudge and an educational intervention. Since our research assistant required unambiguous instructions regarding the selection of studies, we classified any informational intervention as a nudge, including informational programs and workshops, as in Mohammed et al. (2016).

We explicitly instructed our research team to not count financial incentives, legal mandates, medication, and therapy sessions as nudges.

4.2 Outcome Variables

We made the following choices regarding the outcome variables. For smoking we included only papers that addressed cigarette usage. The most commonly used outcome of interest in studies is the probability to *stop* smoking, i.e. a form of extensive margin. For the purposes of our meta-analysis, we therefore focus primarily on this outcome. However, we also have a number of studies that document both the cessation probability *and* changes in number of cigarettes smoked (i.e., the intensive margin) for each subject. We use this information to estimate how the cessation elasticity maps into changes in aggregate cigarette demand.

For influenza vaccinations, our outcome of interest is the vaccination probability. We included only papers that address seasonal flu vaccination, not other epidemic strains such as H1N1.

For electricity consumption, we included all papers that measure residential electricity use. Papers reported outcomes in either absolute terms (kilowatt-hours) or in relative terms (percent

change).

To be able to compare the effects of nudges to price elasticities, we needed to convert the absolute treatment effects into relative percent changes, and transform standard errors of the estimates. Certain information had to be reported in a paper for this to be possible. For the cigarette and vaccine market, this was straightforward since the outcome variables were binary, and means and sample sizes describe entirely the distribution of a binary variable. For energy nudges and price elasticities, this transformation required additional information and we excluded papers that did not report sufficient information. This comprises papers that report absolute treatment effects but no control group consumption, or studies for which no standard errors in percentage terms could be recovered. Generally, estimates of nudge effectiveness and price elasticity had to include either a standard error or sufficient information to impute a standard error. We excluded any paper where such imputation was impossible.

4.3 Data Collection

4.3.1 Scraping

Our data on nudges were obtained chiefly by scraping Microsoft Academic search results using various keywords in the fall of 2020.[4] We chose Microsoft Academic because it was easy to scrape and had similarly good coverage as others potential sources (Martín-Martín et al. 2021). Through our procedure, obtained titles, reference information, abstracts, and (intermittently) lists of citations and references.

After scraping Microsoft Academic, the first 500 results for each search term were checked for relevance by two research assistants working independently, first by title, then abstract. We reconciled the works of the research assistants by checking them against each other, breaking the tie in the event of a disagreement.

This process left us with an abundance of studies, especially for vaccine and cigarette nudges. Two research assistants then read each paper and, if it was deemed fit for inclusion, independently

[4] Unfortunately, Microsoft Academic was discontinued in May 2021.

collected the relevant information – treatment effects, sample sizes, p-values and other statistics, as well as information on the details of the intervention. Afterwards a coauthor reconciled their work, checking them against each other and the paper if there was disagreement.

4.3.2 Leveraging Other Meta-Analyses

The overwhelming majority of papers related to nudging influenza and smoking are from the medical literature. Because the digitization rate of the medical literature, including older literature is fairly good, we were confident in our ability to obtain the relevant sample of papers in these respective literatures. For the literature on energy nudges, we found that this scraping approach did not include some relevant papers. We therefore consulted an additional meta-analysis by Delmas, Fischlein, and Asensio (2013) as well as another survey of the literature by Darby et al. (2006). We searched through the studies that were cited or mentioned in these two additional sources, and adding any that were missed from scraping. Through this procedure, we were able to find 10 additional papers. Of these additional papers, most were old (dating from the first wave of research into behavioral energy reduction following the second oil embargo), grey literature, or had been improperly rejected in an earlier stage of data collection.

There are a variety of high-quality meta-analyses estimating the price elasticities. For cigarette price elasticities, we obtained the underyling data of the meta analysis by Gallet and List (2003). Similarly for energy price elasticities, we collected all the papers mentioned in Zhu et al. (2018) and excluded all the estimates of price elasticity that did not meet our inclusion criteria.

We did not find a meta-analysis for price elasticities of influenza vaccines. To obtain estimates of price elasticity, we searched on Microsoft Academic and Google Scholar for all papers estimating the effects of monetary incentives on vaccination takeup. We also estimated price elasticities from policies that made flu shots free to subjects. Overall, we found substantially more difficult to find studies on price interventions in the market for vaccines than in the other markets.

5 Empirical Results

5.1 Cigarette Consumption

We start by plotting the distribution of nudge and price elasticities. Figure 1 plots a histogram of point estimates of nudge effects together with the estimated log-normal density function. Positive values indicate by how many percent the nudge *increased* the probability of smoking cessation. The underlying data includes 53 point estimates of the effect of nudges on smoking cessation. Each point estimate represents a different study.

The mean of the point estimates is an increase in the cessation probability by 7.5% with a standard deviation of 1.6%. Relative to control, this corresponds to an average increase in the likelihood to quit by 5.6 percentage points. The effect sizes are fairly representative of typical nudge intervention studies in other contexts. The meta-analysis by DellaVigna and Linos (2022) finds an average take-up effect of 8.7 percentage points for academic studies and 1.4 percentage points for studies implemented by "nudge units."

The empirical distribution is left-skewed and has a fairly large standard deviation of 7.5%. The estimated log-normal density provides a good fit of this data. By construction, it has the same mean and standard deviation as stated above, a median of 5.3%, and mode of 2.7%. While most treatment effects are below 10%, the distribution has a long right tail. Since its support is restricted to positive values, it cannot predict the few exceptions in which the treatment effect goes into the "unintended" direction—i.e. nudges that would increase smoking. None of the negative point estimates is statistically different from zero. We provide complementary plots of ranked treatment effects with standard errors in Appendix B.

The distribution of price elasticities of cigarette demand is plotted in Figure 2 and based on 91 point estimates. Positive values indicate a negative price elasticity. The empirical distribution of point estimates has a mean of 0.48 with a standard deviation of 0.19. The standard error of the mean is 0.03, i.e. the average price elasticity is highly statistically different from zero. The mean implies that, on average, a 10% increase in price reduces cigarette demand by 4.8%. The log-

Figure 1: Distribution of Nudge Effects on Cigarette Consumption

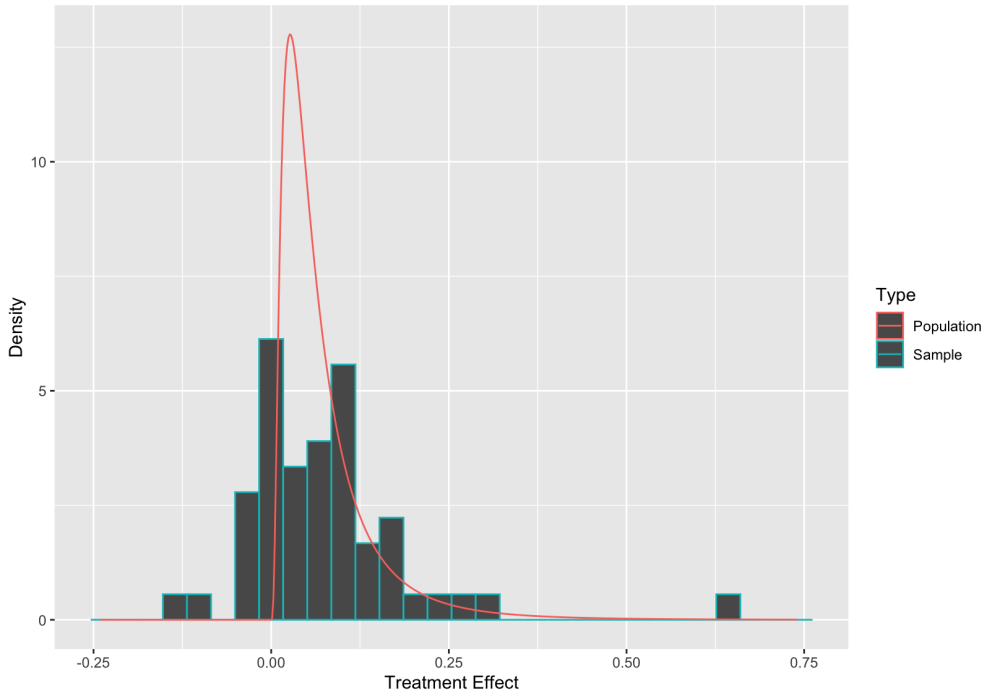
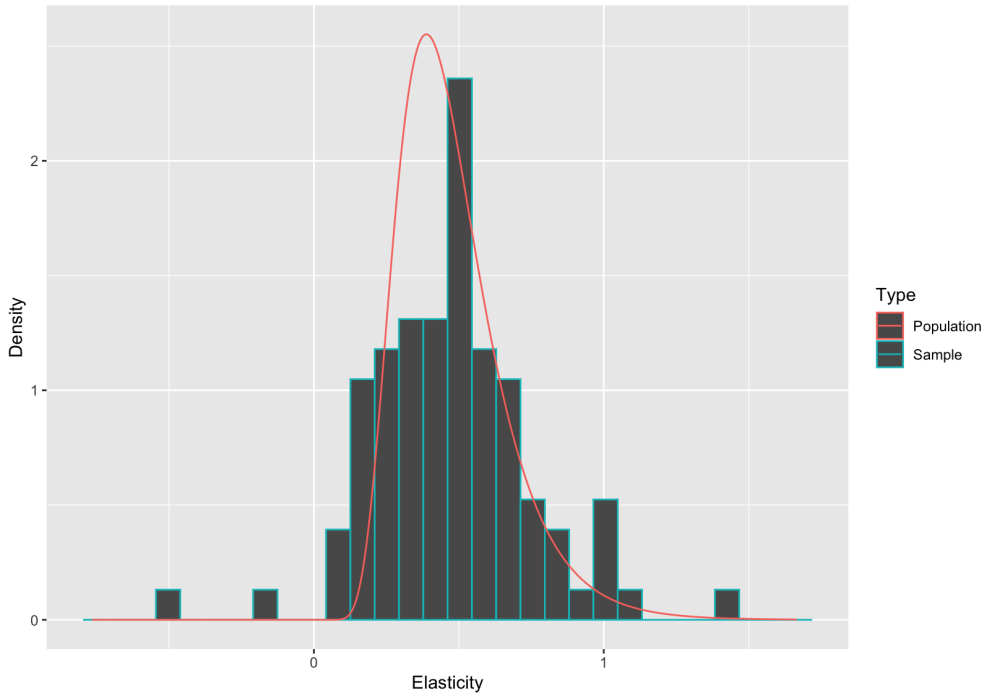


Figure 2: Distribution of Cigarette Price Elasticities



normal population density, again, does a good job at fitting the data. Mean and standard deviation are identical to the distribution of point estimates, with a median of 0.45 and a mode of 0.39.

A first step towards implementing our welfare formulae is understanding how much we would have to tax cigarettes to generate the same effect on aggregate demand as the average nudge does. The challenge in answering this question is that most nudge studies measure the intervention effect on a binary outcome, i.e. cessation, while price studies measure the effect on cigarettes demanded. As a consequence, we need to transform the effect on the cessation probability into cigarette demand. We do this by exploiting eight nudge studies that measure the effect on both the cessation probability and cigarette consumption. We find that a 1% increase in the cessation probability is associated with a 1.8% decrease in cigarette demand. Based on this data we assume that the 7.5% reduction in the cessation probability induced by a nudge implies a 10.26% reduction in cigarette demand. Given our elasticity data, this would mean that prices have to increase by around 20% to induce the same demand reduction as the average nudge does.

The collection of these reduced-form effects is of interest in itself and offers new insights to the literature on nudges and taxes. However, they are not informative about the efficiency effects of these different policy tools. Our theoretical framework allows us to move from reduced-form effects to the quantification of welfare effects. Using equations in 18, we estimate the welfare effects of a nudge, an optimal tax, and a policy mix that uses both the nudge and an optimal tax in combination. We obtain a value for the external damage of smoking from the comprehensive analysis by Sloan et al. (2004) that reports an external damage of 1.44 USD per cigarette pack.[5] Figure 3 plots these effects for different values of the correlation, ρ, between price elasticities and nudge treatment effects. Under quasi-linear utility, this correlation effectively measures how marginal utility of consumption varies with the nudge effect. We further assume that the nudge eliminates 80% of the bias, i.e., $\theta = 0.8$. As we show in the next graph, our qualitative results are not sensitive to this assumption. The y-axis measures gains from each policy in USD per consumer per year. Solid lines represent welfare effects while dashed lines are confidence intervals.

If preferences and nudge effects are independent, $\rho = 0$, the nudge increases welfare by around 0.60 USD per consumer annually. This effect is statistically different from zero. At the same cor-

[5] We assume that one pack has 26 cigarettes.

relation, the point estimate of the optimal tax is only half as a large. In fact, we see that for any value of the correlation, the point estimates suggest that the nudge dominates the tax. A caveat to this conclusion is that there is large uncertainty in the point estimates, as reflected by the large 95% confidence bands.

Another important result to note is that the gains from nudging are decreasing in ρ. To understand this result, first recall that the nudge treatment effect measures by how much consumption deviates from the privately optimal level of consumption. If the treatment effect is larger for more price-elastic consumers, this means that the deviation is larger for consumers with a low marginal utility of consumption. Thus, the behavioral bias distorts consumption where it causes the lowest reduction in consumer surplus. Consumers that are price-inelastic, on the other hand, have smaller treatment effects. Thus, while their welfare-loss per unit of consumption is large, the deviation in consumption from the private optimum is relatively low, such that the welfare loss is small. Conversely, if treatment effect and price elasticity are negatively correlated, the deviation in consumption from the private optimum is larger for consumers with high marginal utility of consumption. The bias then distorts behavior where a unit of consumption is most valued. Since we assume that the nudge can mitigate this heterogeneity, the benefits from nudging are larger (smaller) when the correlation is negative (positive). This result is quantitatively important. The gains from nudging increase by more than 100% as we change the correlation from 1 to -1.

Gains from cigarette taxation are flat in the correlation. This directly follows from our discussion in Section 2.1.3. The behavioral bias only affects gains from taxation through the average marginal bias, $\mathbb{E}_W[b_0]$ and we showed that this statistic is independent of the correlation between treatment effects and price elasticities: $\mathbb{E}_W[b_0] = \frac{TE}{\mathbb{E}[D']}$. While gains from taxation are, in fact, dependent on the correlation between D' and b, as shown in equation 14, this does not mean that they are dependent on the correlation between TE and ϵ. Technically, the reason is that varying $\mathrm{Cov}(TE, \epsilon)$ also mechanically varies the *unweighted* average bias in the population, $\mathbb{E}[b_0]$. These two effects exactly offset each other, such that the average marginal bias, $\mathbb{E}_W[b_0]$, in the population stays the same. As a result, the benefits of taxation are completely independent of the correlation between

nudge effects and price elasticities.

One takeaway from this result is that taxes offer an informational benefit over nudges for policymakers. The evaluation of optimal taxes only requires information on average treatment effects of nudges and prices, while the evaluation of nudges also requires information on the correlation between the two.

The policy mix that combines a nudge and a tax yields slightly larger welfare gains than the nudge in isolation. While the policy mix dominates other policies by construction of our theoretical framework, the magnitudes of the welfare effects are not predetermined by the theory. It is, therefore, interesting to see that a policy mix can barely beat an isolated nudge in the market for cigarettes. Which factors drive this remarkable result? As shown in equation 15, a nudge is particularly powerful relative to a tax if the standard deviation of the bias is larger than the expected externality of consumption. In economic terms, a nudge dominates the tax when targeting of biases is more important than correcting externalities. In the case of smoking, the expected externality is around 1.44 USD per cigarette pack, while the standard deviation of the bias is 1.96. We calculate the ratio of bias standard deviation to average externality as a useful measure for studying the relative benefits of nudging across markets. We refer to this ratio as the *targeting ratio of the nudge*. In the market for cigarettes, this ratio is 1.36 and highly statitstically significantly different from zero, with a bootstraped standard error of 0.41. Thus, the dominating market failure in the context of cigarette consumption is heterogeneity in bias rather than external damages of consumption. This explains why the nudge dominates the tax and why adding a tax to a nudge in a policy mix only provides a small incremental increase in welfare.

So far, we have assumed a nudge effectiveness of 80%. We study the sensitivity of our results to this assumption in Figure 4. The graph plots welfare effects of the policies for different values of nudge effectiveness, holding the correlation parameter fixed at a value of 0.5. We choose a positive correlation because it is most likely to apply if behavioral distortions are smaller for consumers with higher valuations. This assumption follows from models in which consumers optimize with cognitive constraints, such as in models of rational inattention. However, a positive correlation is

a much milder criterion than rational inattention: consumers do not need to allocate attention fully rationally; they only need to be more attentive, on average, when mistakes are more costly.

Figure 3: Welfare Effects in the Cigarette Market for a Nudge Effectiveness of 80%

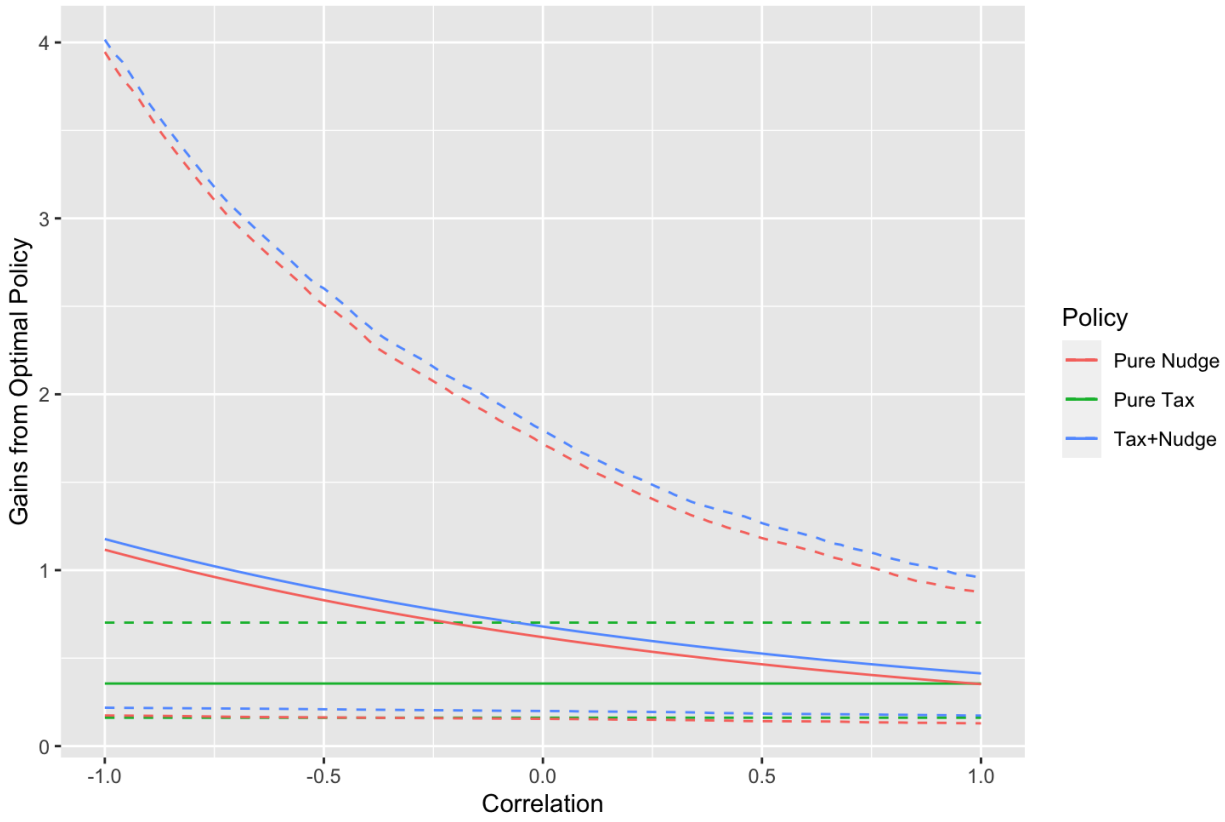
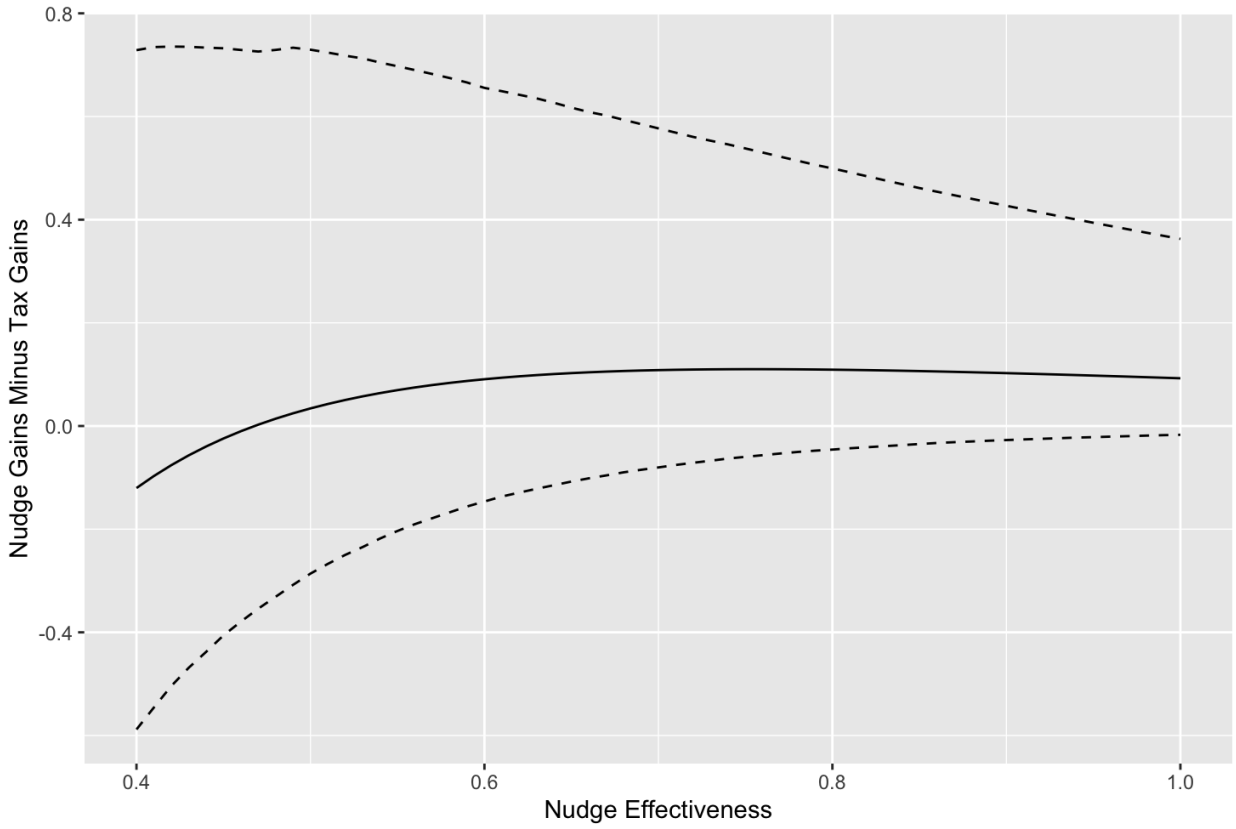


Figure 4 illustrates that the gains from any policy are decreasing in nudge effectiveness. To understand this pattern, recall how the behavioral bias is identified in our model (and in the literature more generally). The magnitude of the behavioral bias is always increasing (linearly) in the nudge treatment effect. This implies that the bias is larger if the nudge only imperfectly de-biases consumers than if it were fully de-biasing. For example, if a nudge reduces the deviation in consumption from the private optimum by 50%, then this implies that the distortion caused by the behavioral bias is twice as large as the treatment effect. Lower values of nudge effectiveness, therefore, imply that the magnitude of the behavioral bias is larger.

In the case of taxation, this mechanically implies that the gains from a corrective tax are larger when nudge effectiveness is lower. In the case of nudging, the directional effect of a change in

Figure 4: Welfare Effects in the Cigarette Market for a Correlation of 0.5



nudge effectiveness is ambiguous. One the one hand, lower nudge effectiveness implies a larger bias, which increases the benefits of nudging. On the other hand, lower nudge effectiveness implies that the nudge is less powerful at de-biasing, which decreases its benefits. Interestingly, our results suggest that the former effect dominates the latter in the case of cigarette consumption.

We find that for most values of nudge effectiveness, the point estimate of the nudge is above the one of the tax. Only when the nudge becomes very ineffective, at $\theta \approx 0.5$, does the tax slightly dominate the nudge. At this level, the nudge and the tax both increase welfare by around 1 USD. The optimal policy mix raises welfare by more than 1.50 USD. As we increase the assumed nudge effectiveness, the differences in benefits between policies shrinks. If we use the usual assumption in the behavioral public economics literature that the nudge is fully-debiasing, the isolated nudge and the policy mix generate virtually the same welfare gains of around 0.40 USD. The isolated tax has a smaller welfare gain of 0.25 USD.

In sum, we find that the potential for welfare gains in the market for cigarettes is larger for nudging than for taxation. This result is robust to a wide range of parameter values for the correlation between preferences and biases, as well as for most levels of nudge effectiveness. The underlying mechanism that generates our results is that heterogeneity in the behavioral bias is more important—from an efficiency perspective—than the external damages from smoking.

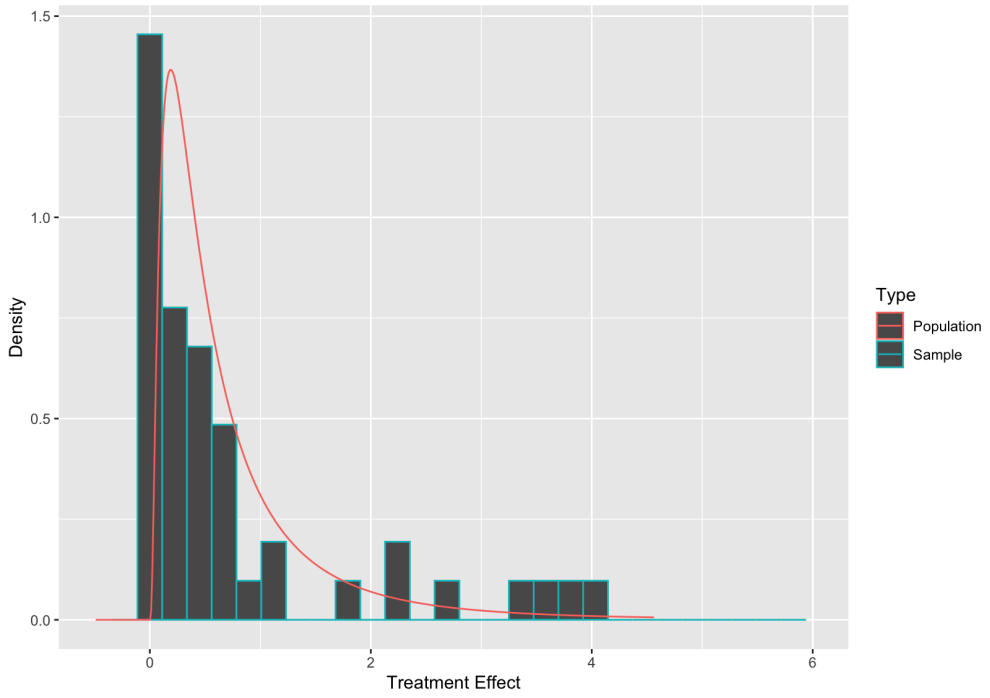
5.2 Influenza Vaccinations

Next, we analyze the effects of nudges and taxes on the take-up of influenza vaccinations. In total, we have 49 studies on the effects of nudges on vaccination take-up. Figure 5 plots the histogram of nudge treatment effects and the population density functions. The mean point estimate is a 1.43% increase in the likelihood to get vaccinated, with a standard error of 0.44%. The standard deviation of the distribution is 2.93%. The vast majority of treatment effects is below 2%. With an average baseline vaccination probability of 31% in the study samples, the mean effect corresponds to an increase by 0.05 percentage points. This estimate is much smaller than in the case of cigarette consumption, but highly statistically significant.

The population density provides a good fit of the histogram with same mean and standard deviation. The median and mode are respectively 0.48% and 0.19%.

While we were able to collect a large number of point estimates on nudge effects, we found that there are surprisingly little studies on the effects of price incentives on vaccination take-up. Most studies focus on behavioral interventions, such as information provision, but rarely implement subsidies that incentivize take-up. One reason for this lack of price studies might be that vaccines are a highly emotionalized topic. Offering money to get vaccinated might be considered morally reprehensible and could eventually backfire. However, among the point estimates we found, there was not a single study that suggested monetary incentives to backfire in the aggregate. Price reductions always increased vaccination take-up. This conclusion is based on a small number of point estimates from 9 studies. The histogram of price elasticities in Figure 6 features a mean absolute elasticity of 0.33 and a standard deviation of 0.35. The mean treatment effect is statistically different

Figure 5: Distribution of Nudge Effects on Vaccination Take-Up



from zero with a standard error of 0.12.

Given the low number of point estimates and a wide dispersion, the population density can only imperfectly capute the data.

In order to generate the same demand response as the average nudge (+1.43%), subsidies would have to reduce the price of influenza vaccines by 4.8%. With a typical price of 41 USD for a standard influenza vaccination, this would imply subsidies of 1.97 USD. Of course, this is not the optimal subsidy level as it does not account for the positive externalities of getting vaccinated.

For our welfare calculations, we assume a positive marginal externality of 153 USD per flu shot. This number is based on the study by White (2021) who estimates an interval of 63 and 243 USD per vaccination. We chose to use the midpoint of this interval. One limitation of our model is that it assumes a constant marginal externality, while the marginal externality of a vaccines is likely to fall with the level of vaccinated people in the population. The optimal Pigouvian subsidy would then need to take on a nonlinear form, which would complicate the analysis and goes beyond the scope of our exercise. In fact, most real-world subsidies on influenza vaccines take a linear form.

Figure 6: Distribution of Vaccination Price Elasticity

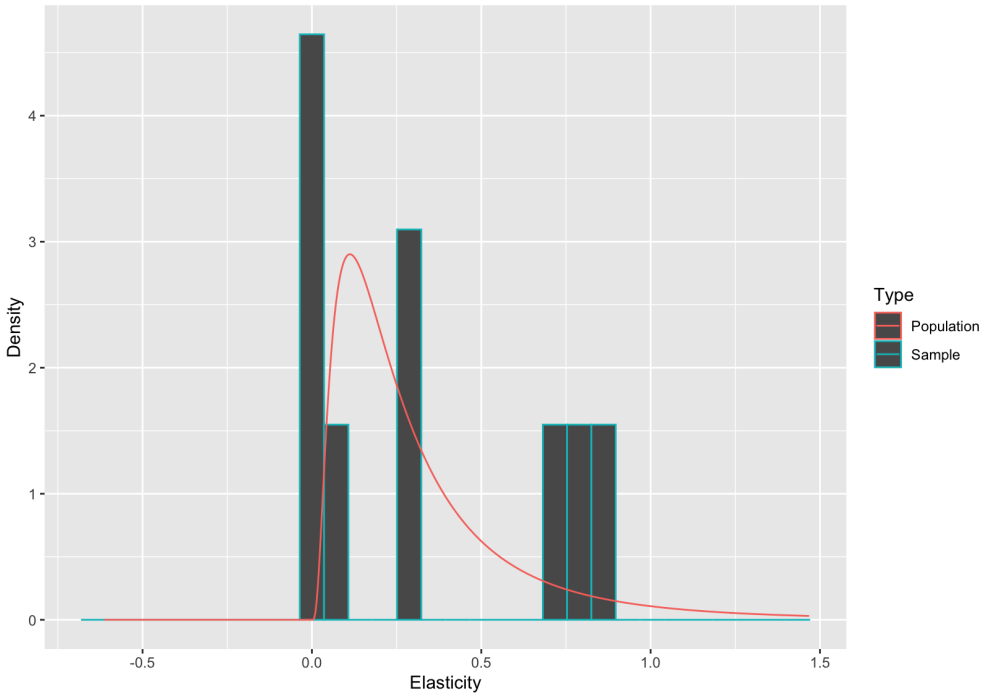


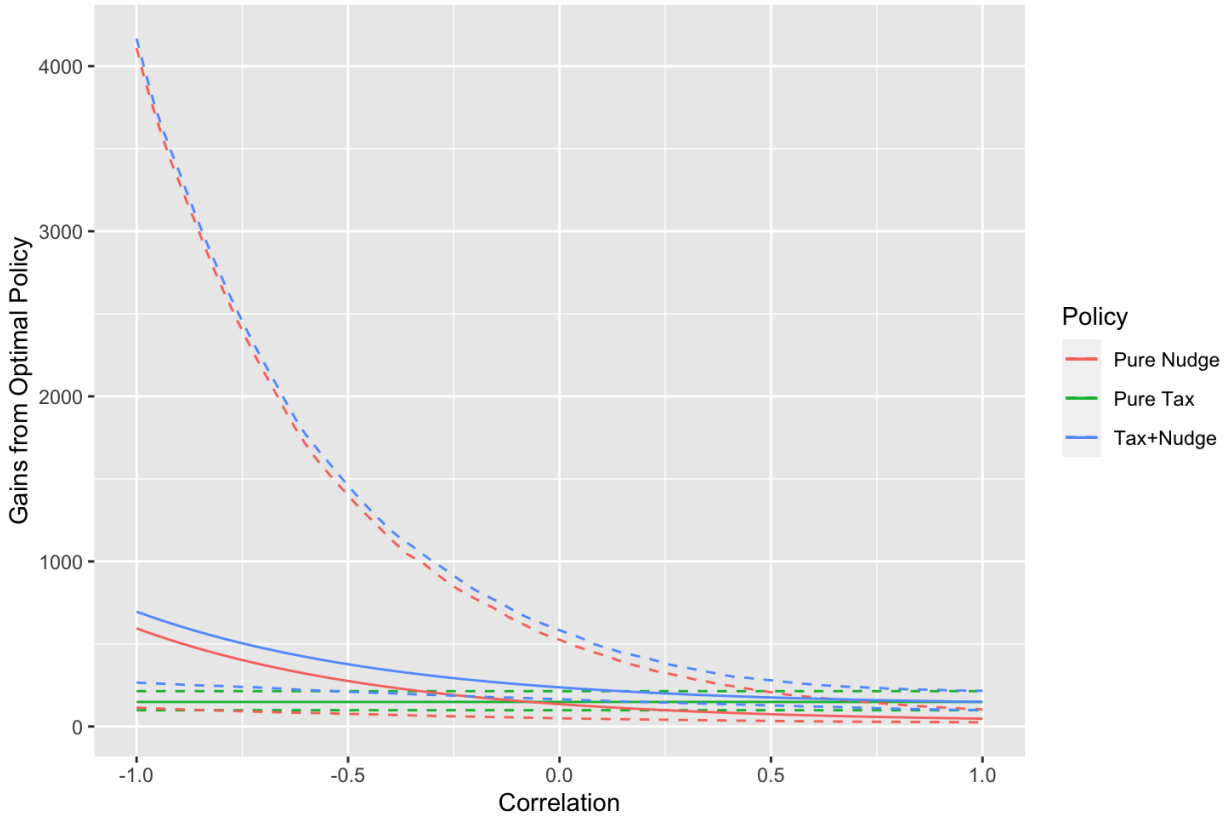
Figure 7 reports welfare effects for a given nudge effectiveness of 80%. The first noticeable observation is that while welfare effects are potentially large, there is extreme uncertainty in the magnitudes. This uncertainty is mostly driven in the low number of price elasticities, which increases standard errors of all three point estimates.

If the correlation between nudge effects and price elasticities is very negative, differences in point estimates becomes large. For a correlation of -1, gains from nudging are estimated to be above 500 USD per consumer, while gains from the optimal subsidy (a negative tax) are half as large. The policy mix only does slightly better than the nudge.

If nudge effects and price elasticities are independent, the gains from nudges and taxes break even. The absolute size of welfare effects becomes substantially smaller (around 200 USD per consumer) as the correlation becomes positive. For positive correlations, point estimates suggest that the optimal subsidy starts dominating the nudge. The intutition for this pattern is the same as in the cigarette market: Positive correlations imply that the behavioral bias causes a lower distortion because large treatment effects are concentrated around consumers with low marginal utility. This

reduces the benefits of targeting the bias, which is precisely the comparative advantage of nudging in our framework. In the extreme case in which the correlation is 1, the benefits of nudging go towards zero and the isolated tax almost performs as well as the policy mix.

Figure 7: Welfare Effects in the Market for Flu Shots, for a Nudge Effectiveness of 80%

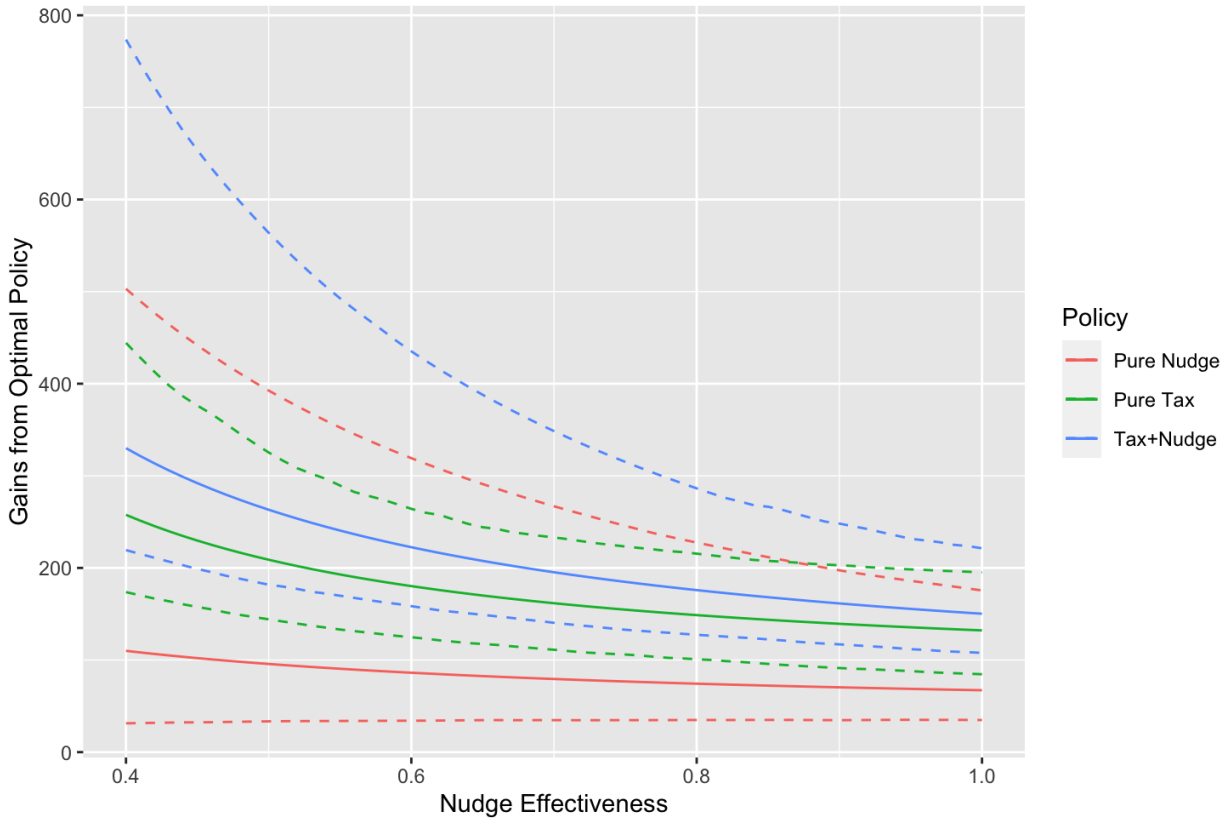


For the likely case in which nudge effects and price elasticities are positively correlated, taxes dominate nudges. Figure 8 illustrates that this statement is true independent of the nudge effectiveness. As in the case of smoking, we plot welfare effects as a function of nudge effectiveness, while fixing the correlation at 0.5. The benefits from the optimal subsidy decrease from 230 USD to 130 USD as we vary nudge effectiveness from 40% to 100%. The effect is unambiguously decreasing because a larger nudge effectiveness implies a lower behavioral bias, which reduces the benefits of a corrective subsidy. The gains from nudging are always strictly below the gains from subsidizing. With lower benefits of nudging, the advantage of having a policy mix over an isolated tax falls, as well.

In conclusion, we obtain quite a different picture than in the market for cigarettes. First, one issue is the substantial uncertainty in welfare effects due to a lack of available price elasticities. This calls for more research on studies examining effects of subsidies on vaccine take-up. Second, if we take the point estimates as our "best guess," the benefits of nudges over subsidies are limited. Nudges are only likely to dominate subsidies if the correlation between preferences and nudge effects is very negative. Further, lower levels of nudge effectiveness increase the benefits of subsidies over nudges.

If we calculate the ratio of standard deviation of bias over average externality, we arrive at a value of 0.27 with a standard error of 1.79. The point estimate suggest that the average externality dominates the heterogeneity in the behavioral bias such that the subsidy would be the preferred instrument. However, the effect is too noisy to draw definitive conclusions.

Figure 8: Welfare Effects in the Market for Flu Shots, for a Correlation of 0.5



5.3 Energy Consumption

To study behaviorally-motivated policies in the energy market, we collect data on energy conservation nudges and energy price elasticities. We focus on nudges that directly affect the end-use of energy in the household. Examples include social comparison nudges that provide households with information on the energy consumption of their neighbors, as well as real-time feedback on people's energy use. We restrict our analysis to electricity consumption because this is the domain for which we found the largest number of nudge- and price-intervention studies.

In total, we collect point estimates from 42 nudge studies whose distribution we plot in Figure 9. The mean treatment effect is a reduction in electricity use by 5.3% with a standard error of 1.3%. The standard deviation of point estimates is 4.5%. There are 5 noisy point estimates that have the unintended sign, i.e. suggest an increase in electricity consumption. None of these estimates is statistically different from zero.

Most nudge interventions have treatment effects below 10% and the median treatment effect is around 4%. The log-normal population density features the same mean, standard deviation and median, and, overall, provides a good fit of the data.

The distribution of price elasticities is shown in Figure 10 and based on 67 studies. The mean elasticity is 0.43 with a tight standard error of 0.04. Thus, a 10% increase in the electricity price causes an average reduction in electricity consumption by 4.3%. This estimate is close to estimates from other meta analyses on electricity price elasticities, such as Labandeira, Labeaga, and López-Otero (2017) who estimate a (absolute) long-term elasticity of 0.53. Our estimates suggest that nudges have the same average effect on demand as a tax of 12.3% on the electricity price.

Almost all price elasticities are smaller than 1 in absolute terms. The distribution of point estimates is notably well represented by the estimated population density function.

For our welfare estimation, we assume that one kWh of electricity produces 0.95kg of CO_2 and that the social cost of carbon is 181 USD per ton of CO_2. The externality value is estimated by Hänsel et al. (2020) who use a version of the Dynamic Integrated Climate–Economy (DICE) inte-

Figure 9: Distribution of Nudge Effects on Electricity Consumption

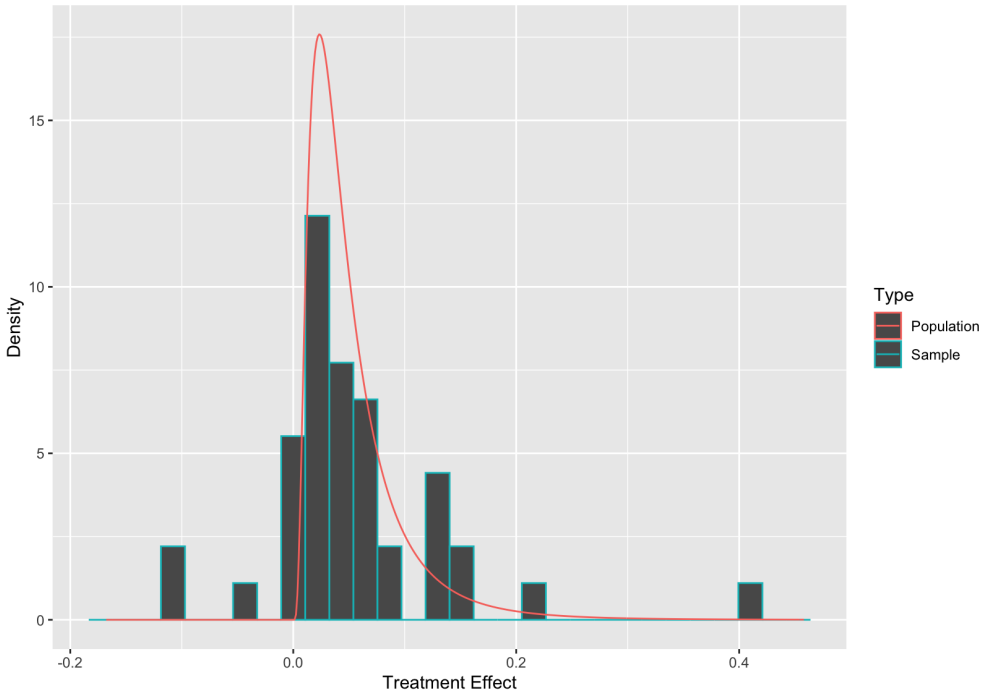
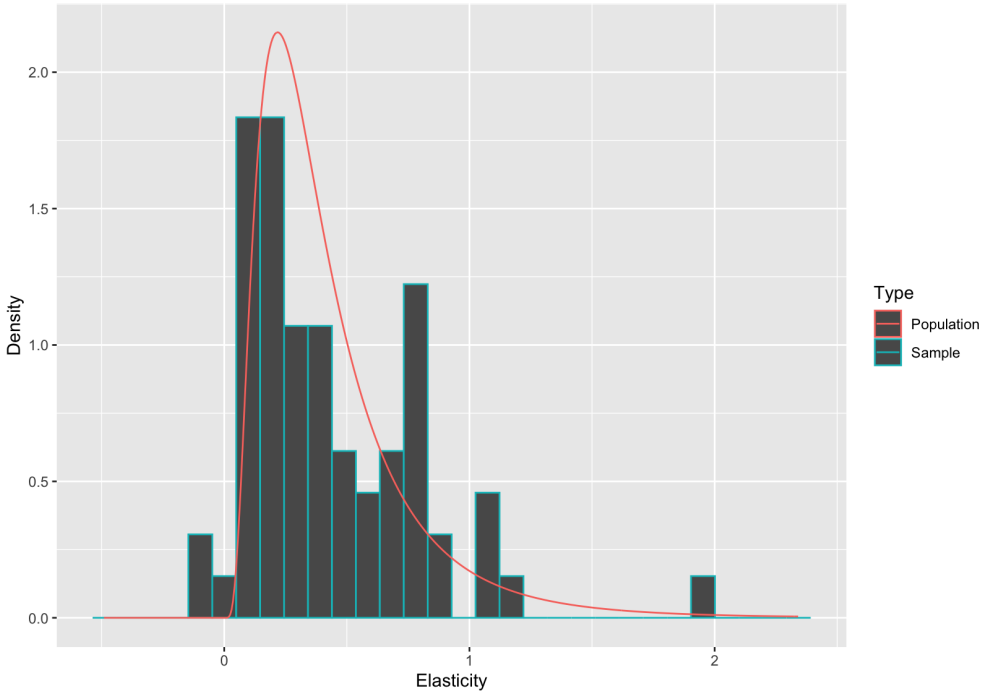


Figure 10: Distribution of Electricity Price Elasticities



grated assessment model together with a wide range of expert views on intergenerational fairness.[6]

[6]Based on the same data, we obtain a standard deviation of the social cost of carbon of 186 USD/tCO_2.

Figure 11 shows welfare effects for our baseline assumption that the nudge eliminates 80% of the behavioral bias. It becomes immediately apparent that regulation in the energy market is fundamentally different from the previous examples. For any correlation between price elasticities and nudge effects, benefits from taxation vastly exceed benefits from nudging. The nudge increases welfare by between \$49 and \$94 per household per year. The optimal tax raises social welfare by close to \$882, i.e., roughly 9-18 times more than the nudge. The difference in welfare gains between the two tools is highly statistically significant. A notable result is that a policy mix of nudges and taxes can barely beat the isolated tax. The only case in which the gains from a policy mix exceed the gains from a isolated tax is in the fairly unlikely situation in which the correlation between preferences and nudge effects is extremely negative. Under no assumptions is the difference in gains between the policy mix and the isolated tax larger than 5%.

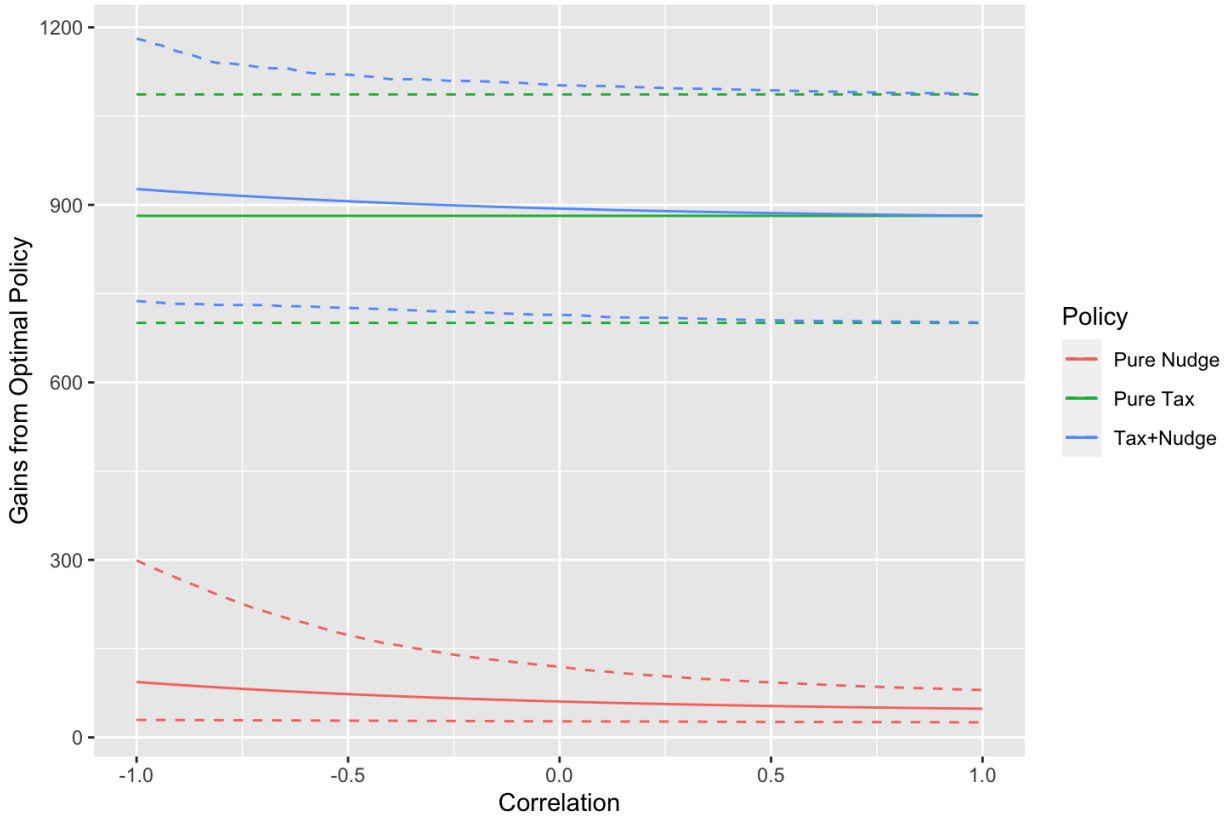
Again, a comparison of the standard deviation of the bias with the average externality can explain these results. The standard deviation of the behavioral bias is \$0.016 while the average marginal externality per kwH is \$0.19. The ratio of these two is 0.05 with a tight standard error of 0.01. Thus, the benefits from externality correction vastly exceed the benefits from targeting the behavioral bias.

Our results remain virtually unchanged for different values of nudge effectiveness and a given value of the correlation of 0.5. As shown in Figure 12, the benefits of taxation still amount to around 900 USD per household for many values of nudge effectiveness. As in the prior examples, lower nudge effectiveness implies that the behavioral bias becomes larger, which increases the benefits of a corrective tax. For very low levels of nudge effectiveness, gains from taxation approach 1,000 USD per household per year.

Gains from nudging remain relatively low, ranging from \$3 to \$63 per consumer. Interestingly, this time gains are slightly increasing in nudge effectiveness. Thus, while there is less need for policy intervention for larger values of nudge effectiveness, gains from nudging are still increasing because the policy tool becomes more effective at correcting the bias. However, even for favorable parameter values of nudge effectiveness, we find that benefits from taxation are 1,038% larger than benefits from nudging. The policy mix has virtually no additional benefit over the optimal tax.

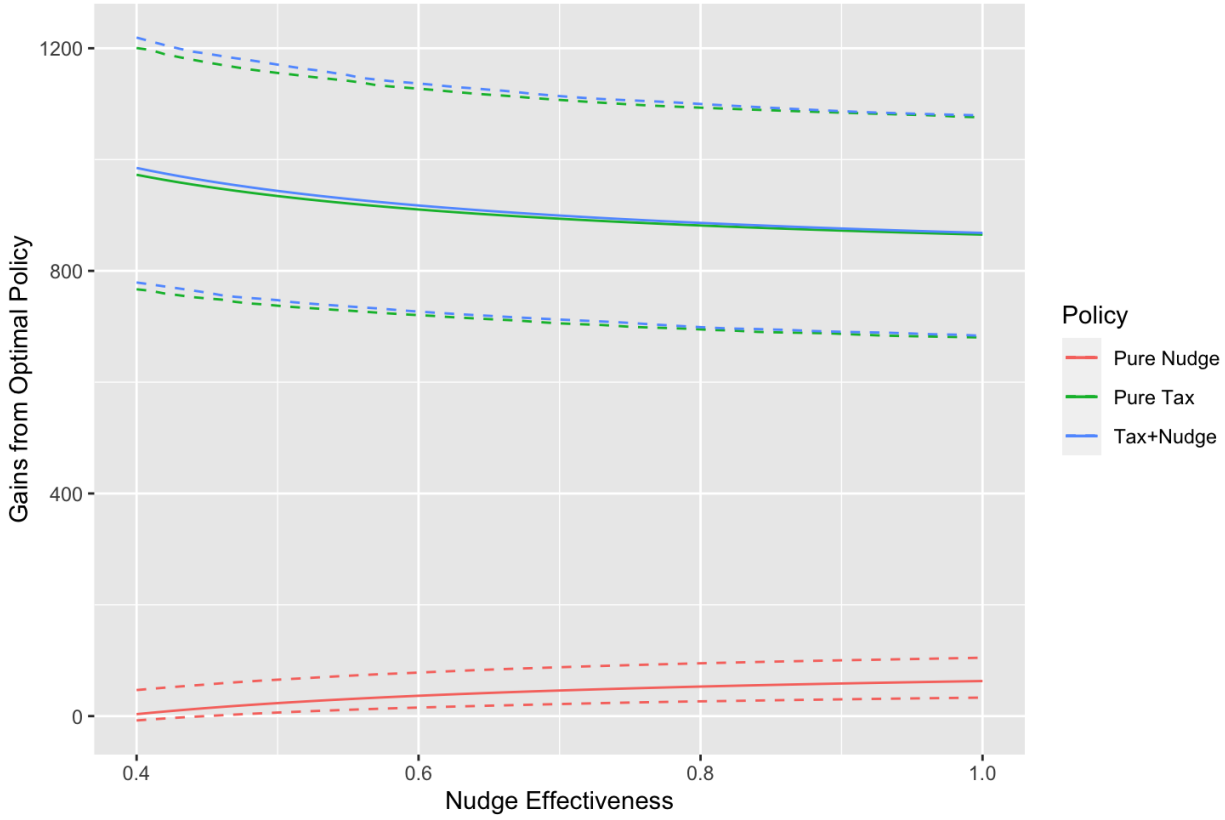
Different from the other markets, the present results suggests that policymakers and academics may want to focus on the implementation of optimal price regulations rather than on behavioral interventions. Since the benefits of the two policies are so extremely different in terms of magnitudes, the opportunity cost of studying and implementing nudges in the energy market appears high. This insight is particularly relevant for current policy debates about optimal environmental policy. Most governments have not yet put a price on carbon, either through a carbon tax or a cap-and-trade system. At the same time, energy conservation nudges have been implemented at an extremely large scale by many governments around the globe. While our results do not imply that these behavioral policies provide no benefits, they do suggest that substantially more attention should be paid towards the political implementation of a carbon price.[7]

Figure 11: Welfare Effects in the Electricity Market for a Nudge Effectiveness of 80%



[7] In fact, behavioral interventions may prove useful in reducing the prevalent opposition against a carbon tax in many countries.

Figure 12: Welfare Effects in the Electricity Market for a Correlation of 0.5



6 Conclusion

What do hundreds of studies on the reduced-form effects of nudges tell us about their welfare effects? How do traditional tax policies compare to these novel behavioral interventions in terms of economic efficiency? We answer these questions by implementing the first meta-analysis on the welfare effects of behavioral public policies. Based on an established theoretical framework in public economics, we derive sufficient statistics of the welfare effects of nudges, taxes, and a policy mix that used both tools. We show that point estimates of nudge and price effects suffice to quantify efficiency effects for given values of the correlation between preferences and behavioral biases.

We implement our results empirically by collecting a large number of point estimates on the effects of nudges and taxes. We apply our framework to three widely regulated settings in which behavioral biases are allegedly ubiquitous. We find that both nudges and price interventions are effective in reducing smoking, in increasing the take up of influenza vaccination, as well as in

conserving household electricity. We then take these reduced-form effects and estimate welfare effects for each intervention, leveraging our theoretical framework.

While nudges are *effective* in changing behavior, this does not imply that they are the more efficient or "low-cost" alternative to taxation. We find that nudges are likely to be more socially efficient in the market for cigarettes, but are far less efficient than taxes in the electricity market. For flu shots, the effects are too noisy to draw definitive conclusions, but subsidies seem to have a slight advantage for reasonable parameter values. We show that two factors govern the difference in results across markets in our framework: i) the standard deviation of the behavioral bias, and ii) the magnitude of the average externality. Nudges have the unique advantage over taxes that they potentially reduce the heterogeneity in the behavioral distortion. Taxes, on the other hand, have the unique advantage of internalizing the externality. Whenever the heterogeneity in bias is large relative to the size of the externality, nudges dominate taxes. This insight highlights the importance for academic studies to estimate these statistics in various markets that are of interest for policymakers. As such, they can provide important guidance for the optimal design of behavioral public policy.

Our theoretical framework implies that a combination of nudges and taxes always outperforms each policy in isolation. However, empirically it turns out that there is large variation in how much combining policy tools add to social welfare. Under certain parameter values, adding a tax/subsidy to a nudge can provide important incremental efficiency gains in the market for cigarettes and influenza vaccines. However, in the electricity market, the additional benefit of adding a nudge to a tax is vanishingly small under virtually all parameter values. This exercise highlights the importance of the empirical quantification of welfare effects.

Finally, our empirical analyses relied on assumptions about the effectiveness of nudges in debiasing consumers, as well as on the correlation between price and nudge effects. While most of our qualitative conclusions are insensitive to these assumptions, quantitative results may change. This calls for more empirical designs that i) understand how well nudges reduces cognitive biases, and ii) implement both nudge and price interventions in the same sample and estimate the covariance

of these effects. Future research can make important contributions in quantifying these parameters.

References

Allcott, Hunt. 2011. "Social norms and energy conservation." *Journal of Public Economics* 95 (9–10): 1082–1095.

Allcott, Hunt, and Judd B. Kessler. 2019. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *American Economic Journal: Applied Economics* 11 (1): 236–76 (January).

Allcott, Hunt, and Christopher Knittel. 2019. "Are Consumers Poorly Informed About Fuel Economy? Evidence from Two Experiments." *American Economic Journal: Economic Policy* 11 (1): 1–37.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019. "Regressive sin taxes, with an application to the optimal soda tax." *The Quarterly Journal of Economics* 134 (3): 1557–1626.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky. 2014. "Energy Policy with Externalities and Internalities." *Journal of Public Economics* 112:72–88.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *The American Economic Review* 104 (10): 3003–3037.

Allcott, Hunt, and Dmitry Taubinsky. 2015. "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market." *The American Economic Review* 105 (8): 2501–2538.

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

Andor, Mark, Andreas Gerster, Jörg Peters, and Christoph M Schmidt. 2018. "Social norms and energy conservation beyond the US." *USAEE Working Paper*.

Antinyan, Armenak, and Zareh Asatryan. 2019. "Nudging for tax compliance: A meta-analysis." Technical Report 19-055.

Armitage, Christopher J, and Madelynne A Arden. 2008. "How useful are the stages of change for targeting interventions? Randomized test of a brief intervention to reduce smoking." *Health Psychology* 27 (6): 789.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C Teitelbaum. 2013. "The nature of risk preferences: Evidence from insurance choices." *American economic review* 103 (6): 2499–2529.

Benartzi, Shlomo, John Beshears, Katherine L Milkman, Cass R Sunstein, Richard H Thaler, Maya Shankar, Will Tucker-Ray, William J Congdon, and Steven Galing. 2017. "Should governments invest more in nudging?" *Psychological science* 28 (8): 1041–1055.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *The Quarterly Journal of Economics* 124 (1): 51–104.

Borland, Ron, James Balmford, and Elena Swift. 2015. "Effects of Encouraging Rapid Implementation and/or Structured Planning of Quit Attempts on Smoking Cessation Outcomes: a Randomized Controlled Trial." *Annals of Behavioral Medicine* 49 (5): 732–742 (04).

Butera, Luigi, Robert Metcalfe, William Morrison, and Dmitry Taubinsky. 2022. "Measuring the welfare effects of shame and pride." *American Economic Review* 112 (1): 122–68.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *The American Economic Review* 99 (4): 1145–1177.

Conlin, Michael, Ted O'Donoghue, and Timothy J Vogelsang. 2007. "Projection bias in catalog orders." *American Economic Review* 97 (4): 1217–1249.

Darby, Sarah, et al. 2006. "The effectiveness of feedback on energy consumption." *A Review for DEFRA of the Literature on Metering, Billing and direct Displays* 486, no. 2006.

DellaVigna, Stefano. 2018. "Structural behavioral economics." In *Handbook of Behavioral Economics: Applications and Foundations 1*, Volume 1, 613–723. Elsevier.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to scale: Comprehensive evidence from two nudge units." *Econometrica* 90 (1): 81–116.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *The Quarterly Journal of Economics* 127 (1): 1–56 (February).

DellaVigna, Stefano, John A. List, Ulrike Malmendier, and Gautam Rao. 2016. "Voting to Tell Others." *The Review of Economic Studies* 84 (1): 143–181.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao. 2022. "Estimating Social Preferences and Gift Exchange at Work." *American Economic Review* 112 (3): 1038–74.

Delmas, Magali A., Miriam Fischlein, and Omar I. Asensio. 2013. "Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012." *Energy Policy* 61:729–739.

Farhi, Emmanuel, and Xavier Gabaix. 2020. "Optimal Taxation with Behavioral Agents." *American Economic Review* 110 (1): 298–336.

Frank, John W, Lorraine McMurray, and Margaret Henderson. 1985. "Influenza vaccination in the elderly: 2. The economics of sending reminder letters." *Canadian Medical Association Journal* 132 (5): 516.

Gallet, Craig A., and John A. List. 2003. "Cigarette demand: a meta-analysis of elasticities." *Health Economics* 12 (10): 821–835.

Hänsel, Martin C, Moritz A Drupp, Daniel JA Johansson, Frikk Nesje, Christian Azar, Mark C

Freeman, Ben Groom, and Thomas Sterner. 2020. "Climate economics support for the UN climate targets." *Nature Climate Change* 10 (8): 781–789.

Harberger, Arnold C. 1964. "The Measurement of Waste." *The American Economic Review* 54 (3): 58–76.

Hedblom, Daniel, Brent R Hickman, and John A List. 2019. "Toward an understanding of corporate social responsibility: Theory and field experimental evidence." Technical Report, National Bureau of Economic Research.

Hennrikus, Deborah J., Harry A. Lando, Maribet C. McCarty, David Klevan, Neal Holtan, Jacquelyn A. Huebsch, Sharon Jestus, Paul R. Pentel, Donald Pine, Susan Sullivan, Karen Swenson, and John Vessey. 2005. "The TEAM project: the effectiveness of smoking cessation intervention with hospital patients." *Preventive Medicine* 40 (3): 249–258.

Higgins, JPT, J Thomas, J Chandler, M Cumpston, T Li, MJ Page, and VA Welch, eds. 2022, Feb. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane.

Houde, Sébastien. 2018. "How consumers respond to product certification and the value of energy information." *The RAND Journal of Economics* 49 (2): 453–477.

Hummel, Dennis, and Alexander Maedche. 2019. "How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies." *Journal of Behavioral and Experimental Economics* 80:47–58.

Jessoe, Katrina, and David Rapson. 2014. "Knowledge is (less) power: Experimental evidence from residential energy use." *American Economic Review* 104 (4): 1417–38.

Labandeira, Xavier, José M Labeaga, and Xiral López-Otero. 2017. "A meta-analysis on the price elasticity of energy demand." *Energy policy* 102:549–568.

Laibson, David, Andrea Repetto, and Jeremy Tobacman. 2007. "Estimating discount functions with consumption choices over the lifecycle." Technical Report.

Löschel, Andreas, Matthias Rodemeier, and Madeline Werthschulte. 2022. "Can Self-Set Goals

Encourage Resource Conservation? Field Experimental Evidence from a Smartphone App." Working Paper.

Martín-Martín, Alberto, Mike Thelwall, Enrique Orduna-Malea, and Emilio Delgado López-Cózar. 2021. "Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: a multidisciplinary comparison of coverage via citations." *Scientometrics* 126 (1): 871–906.

Milkman, Katherine L, John Beshears, James J Choi, David Laibson, and Brigitte C Madrian. 2011. "Using implementation intentions prompts to enhance influenza vaccination rates." *Proceedings of the National Academy of Sciences* 108 (26): 10415–10420.

Mohammed, Mutaz, Sander Matthijs Eggers, Fahad F. Alotaiby, Nanne de Vries, and Hein de Vries. 2016. "Effects of a randomized controlled trial to assess the six-months effects of a school based smoking prevention program in Saudi Arabia." *Preventive Medicine* 90:100–106.

Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon. 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4 (1): 511–540.

Rodemeier, Matthias. 2020. "Buy baits and consumer sophistication: Theory and field evidence from large-scale rebate promotions." CAWM Working Paper 124.

Rodemeier, Matthias, and Andreas Löschel. 2022. "Information Nudges, Subsidies, and Crowding Out of Attention: Field Evidence from Energy Efficiency Investments." Working Paper.

Sloan, Frank A, Jan Ostermann, Christopher Conover, Donald H Taylor Jr, and Gabriel Picone. 2004. *The price of smoking*. MIT press.

Srinivasan, Mythili, Jonica Huntman, Miranda Nelson, and Shakila Mathew. 2020. "Use of Peer Comparison, Provider Education, and Electronic Medical Record Triggers to Increase Influenza Vaccination Rates in Hospitalized Children." *Hospital Pediatrics* 10 (1): 76–83 (01).

Stefanski, Leonard A, and Raymond J Carroll. 1990. "Deconvolving kernel density estimators." *Statistics* 21 (2): 169–184.

White, Corey. 2021. "Measuring social and externality benefits of influenza vaccination." *Journal of Human Resources* 56 (3): 749–785.

Zhu, Xing, Lanlan Li, Kaile Zhou, Xiaoling Zhang, and Shanlin Yang. 2018. "A meta-analysis on the price elasticity and income elasticity of residential electricity demand." *Journal of Cleaner Production* 201:169–177.

A Derivation of Welfare Formulae Under Log-Normal Model

In this appendix, we derive the Welfare Formulae of Equation (18). The following facts about log-normal random variables are used throughout. If $X \sim \log\mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \log\mathcal{N}(\mu_Y, \sigma_Y^2)$, then $\mathbb{E}[X] = \exp(\mu + \sigma^2/2)$ while $\mathrm{Var}(X) = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)$. Additionally, if $\log X, \log Y$ are correlated with correlation coefficient ρ, then $XY \sim \log\mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y)$

A.1 Welfare Loss from Doing Nothing

If nothing is done, welfare loss relative to first best is given by

$$-\frac{1}{2}\left(\mathbb{E}[\xi^2]\mathbb{E}[S] + 2\mathbb{E}[BS]\mathbb{E}[\xi] + \mathbb{E}[B^2S]\right).$$

Since monomials of log-normal random variables are again log-normal, we have that B^2S is lognormal with $\mathbb{E}[B^2S] = \exp(2\mu_B + \mu_S + [4\sigma_B^2 + 4\sigma_{BS} + \sigma_S^2]/2)$. Meanwhile, $\mathbb{E}[\xi^2] = \exp\left(2\mu_\xi + 2\sigma_\xi^2\right)$ and $\mathbb{E}[S] = \exp(\mu_S + \sigma_S^2/2)$. Putting this all together, we have that

$$\begin{aligned} L = \frac{1}{2}\{&\exp\left(2\mu_\xi + \mu_S + [4\sigma_\xi^2 + \sigma_S^2]/2\right) \\ &+ \exp(\mu_B + \mu_S + \mu_\xi + [\sigma_B^2 + 2\sigma_{BS} + \sigma_S^2 + \sigma_\xi^2]/2) \\ &+ \exp\left(2\mu_B + \mu_S + [4\sigma_B^2 + 4\sigma_{BS} + \sigma_S^2]/2\right)\} \end{aligned}$$

A.2 Welfare Loss from Optimal Tax in Isolation

Recall that the welfare formula in this case is given by $-\frac{1}{2}\mathbb{E}[S]\mathrm{Var}_S(b+\xi)$. Under the assumed independence between ξ and S, $\mathrm{Var}_S(\xi) = \mathrm{Var}(\xi) = [\exp(\sigma_\xi^2) - 1]\exp(2\mu_\xi + \sigma_\xi^2)$. On the other hand, by definition, we have that

$$\begin{aligned}
-\mathbb{E}[S]\mathrm{Var}_S[B] &= \mathbb{E}[SB^2] - \frac{1}{\mathbb{E}[S]}\mathbb{E}[SB]^2 \\
&= \exp\left(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2\right) \\
&\quad - \exp\left(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 2\sigma_B^2]/2\right) \geq 0
\end{aligned}$$

with strict inequality whenever $\sigma_B^2 > 0$. Putting this all together, we have that

$$\begin{aligned}
L = \frac{1}{2}\{[\exp(\sigma_\xi^2) - 1]\exp(2\mu_\xi + \mu_S + \sigma_\xi^2 + \sigma_S^2/2) + \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\
- \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 2\sigma_B^2]/2)\}
\end{aligned}$$

A.3 Welfare Loss from Isolated Nudging

Recall that the welfare loss of the partial nudge is given by $-\frac{1}{2}\mathbb{E}[S]\mathbb{E}_S[((1-\theta)B + \xi)^2]$. This formula is equal to

$$(1-\theta)^2\mathbb{E}[SB^2] + 2(1-\theta)\mathbb{E}[SB]\mathbb{E}[\xi] + \mathbb{E}[S]\mathbb{E}[\xi^2]$$

As before, each of the expectations above are of log-normal random variables, so we have

$$\begin{aligned}
L = \frac{1}{2}\{&(1-\theta)^2\exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\
&+2(1-\theta)\exp(\mu_S + \mu_B + \mu_\xi + [\sigma_S^2 + 2\sigma_{BS} + \sigma_B^2 + \sigma_\xi^2]/2) \\
&+\exp(\mu_S + 2\mu_\xi + [\sigma_S^2 + 4\sigma_\xi^2]/2)\}
\end{aligned}$$

A.4 Welfare Loss from Nudging + Optimal Taxation

The welfare formula in this case is given by $-\frac{1}{2}\mathbb{E}[S]\mathrm{Var}((1-\theta)B+\xi)$. Rearranging the calculations already done for the isolated tax case, we have that

$$\begin{aligned} L = \frac{1}{2}\{&[\exp(\sigma_\xi^2) - 1]\exp(2\mu_\xi + \mu_S + \sigma_\xi^2 + \sigma_S^2/2) \\ &+(1-\theta)^2[\exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 4\sigma_B^2]/2) \\ &\quad - \exp(\mu_S + 2\mu_B + [\sigma_S^2 + 4\sigma_{BS} + 2\sigma_B^2]/2)]\} \end{aligned}$$

B Point Estimates and Standard Errors for Each Study

Figure 13: Nudge Treatment Effects on Smoking Cessation Probability

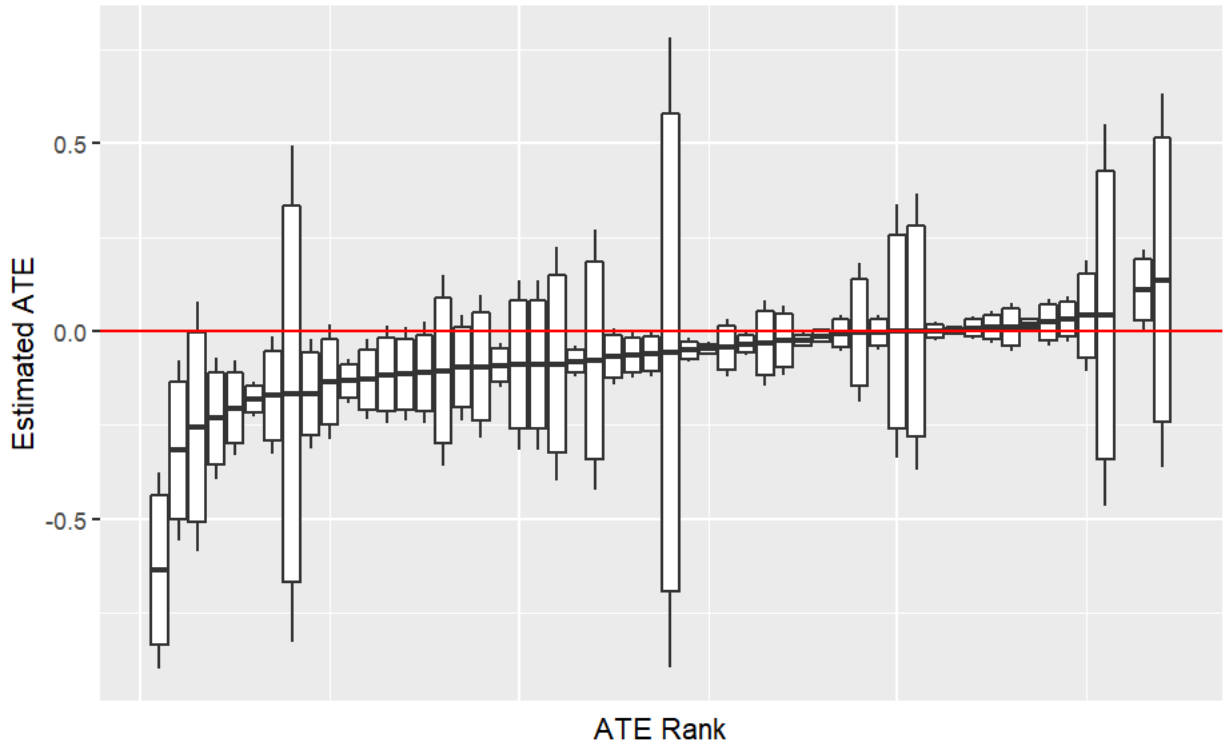


Figure 14: Cigarette Price Elasticities

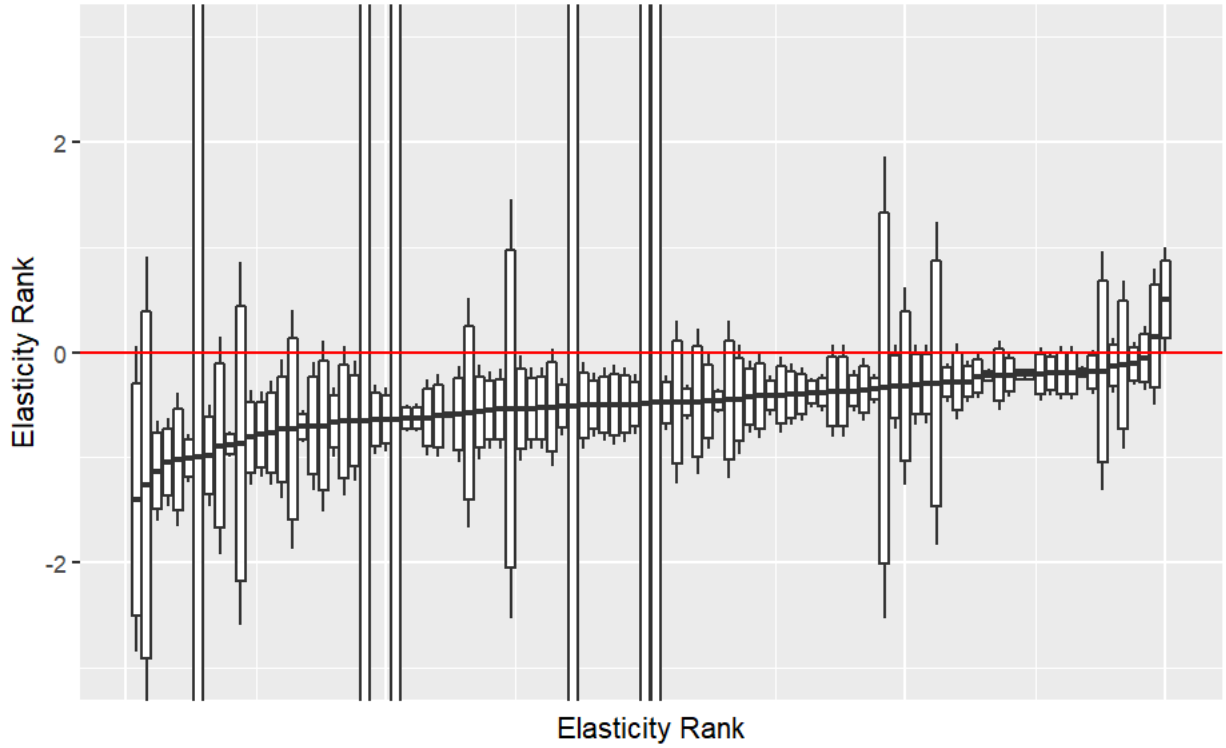


Figure 15: Nudge Treatment Effects on Influenza Vaccination Probability

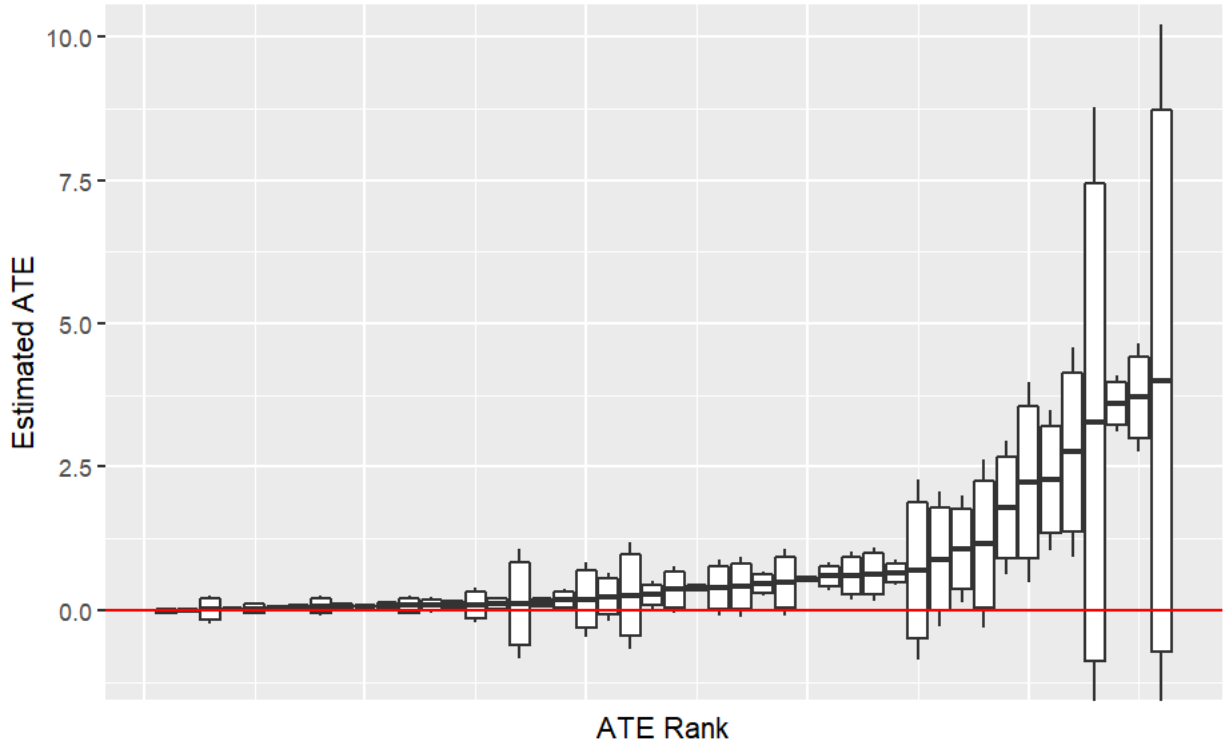


Figure 16: Influenza Vaccine Price Elasticities

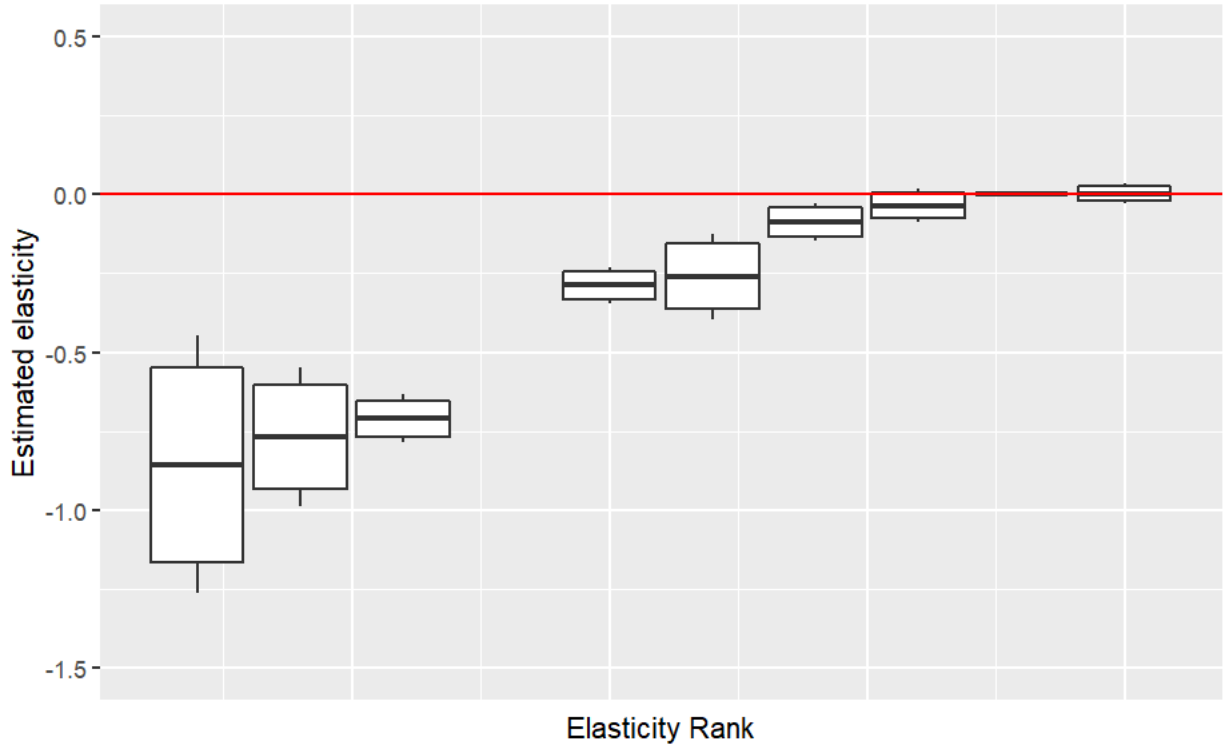


Figure 17: Nudge Treatment Effects on Electricity Consumption

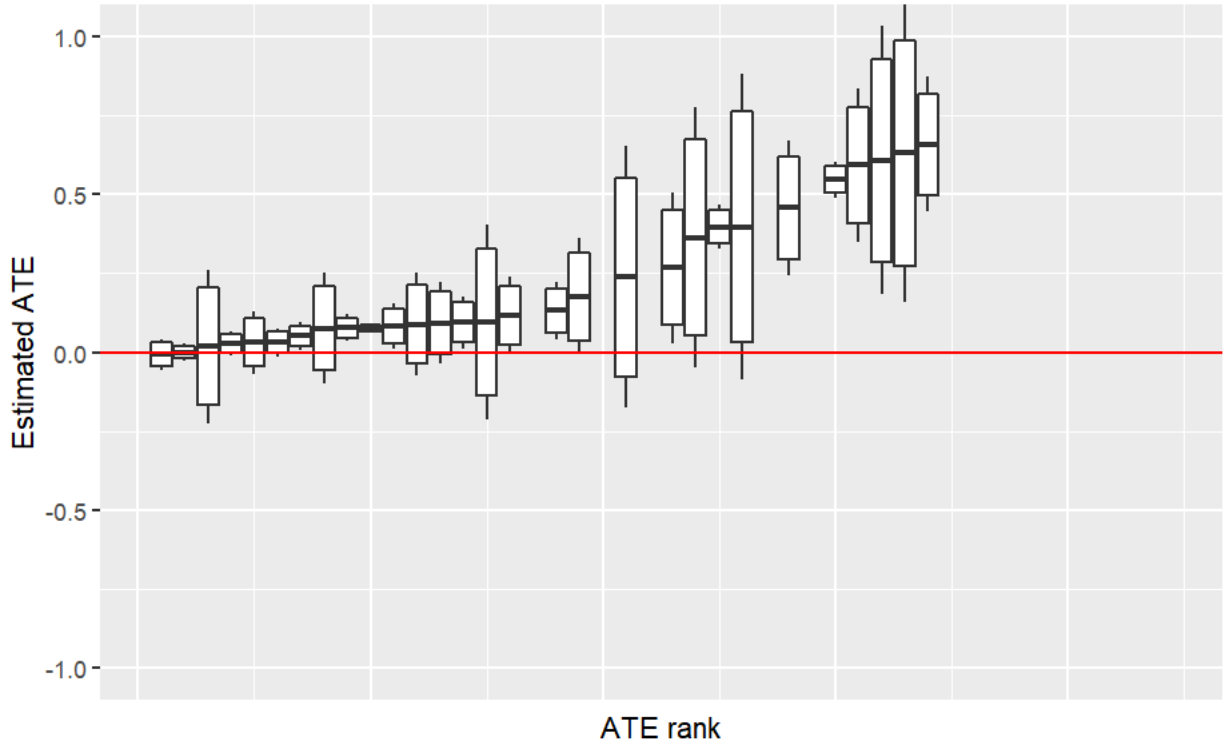
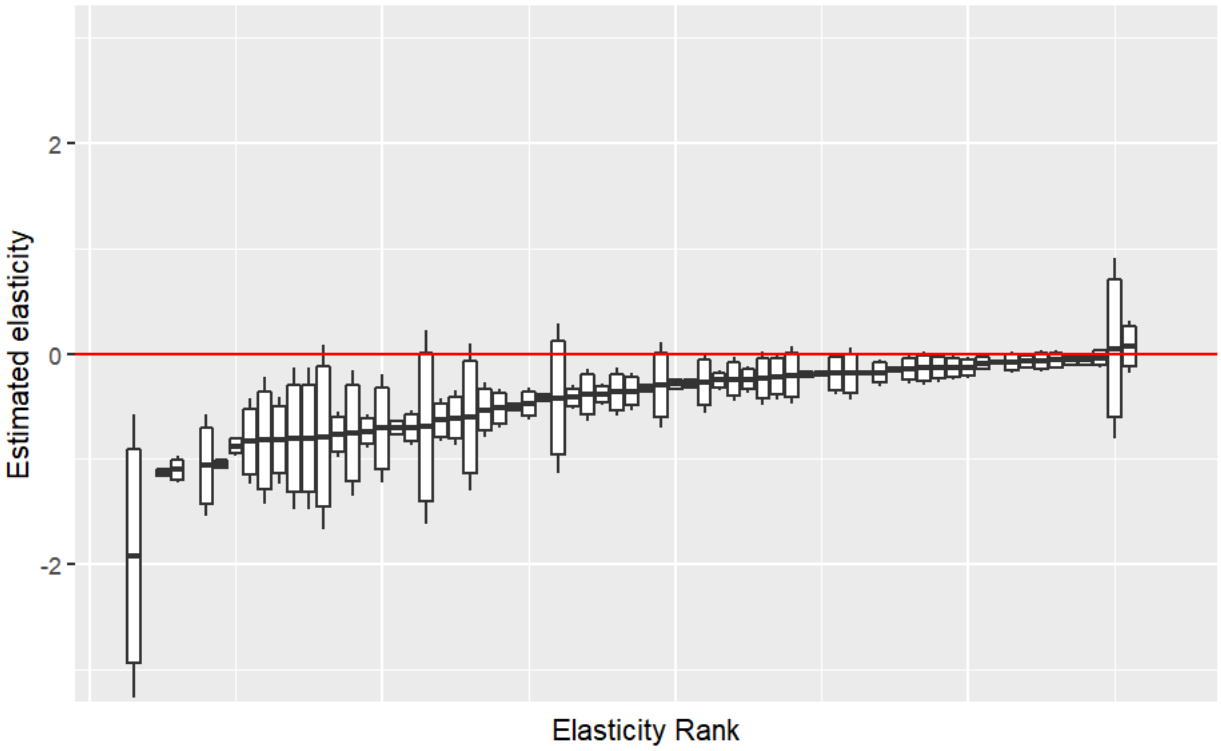


Figure 18: Electricity Price Elasticities



C Structural Parameters and Optimal Taxes

Table 1: Optimal Taxes

	Cigarettes (per pack)	Influenza Vaccines (per vaccine)	Electricity (per kWh)
EPM of behavioral bias	\$5.1 (\$1.1)	-\$75 (\$202)	\$0.02 (\$0.005)
Optimal isolated tax	\$6.5 (\$1.1)	-\$228 (\$202)	\$0.21 (\$0.005)
Optimal tax with Nudge	\$2.5 (\$0.2)	-\$168 (\$40)	\$0.19 (\$0.001)
Targeting ratio of nudge: $SD_b/Externality$	1.36 (0.41)	0.27 (1.79)	0.051 (0.014)

Notes: This table reports the equivalent price metric in each market, as well as the size of the optimal tax with and without nudge. The last row indicates the relative targeting ratio, which is defined as the standard deviation of the behavioral bias over the average externality. Standard errors in parentheses.