ESSAYS IN BEHAVIORAL ECONOMICS AND TAXATION

Jacob Goldin

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF ECONOMICS

Advisors: Henry Farber and David Lee

SEPTEMBER 2015
Copyright by Jacob Goldin, 2015. All rights reserved.
Abstract

This dissertation studies behavioral economics and the design of public policy. Chapters 1 and 2 investigate tax salience — the notion that the prominence of a tax shapes the extent to which taxpayers account for it when making purchasing decisions. In many contexts, policymakers can control a tax’s salience and Chapter 1 investigates how such control should be exercised to best promote consumer welfare. Chapter 2 turns to distributional concerns relating to tax salience. For some taxes, salience effects may vary by income; in work co-authored with Tatiana Homonoff, I investigate how policymakers can take advantage of this fact to make the distribution of such taxes less regressive. Drawing on state and time variation in cigarette tax rates, we empirically investigate whether salience effects vary by income and find evidence consistent with this theory.

Chapter 3 turns away from tax salience to broader issues concerning behavioral economics and public policy. In numerous settings, behavior varies according to seemingly arbitrary features of the decision-making environment, such as which option is the default, the order in which options are presented, or which option characteristics are salient. Optimal policy design requires accounting for the preferences of decision-makers whose choices are sensitive to such factors, but traditional revealed preference analysis breaks down in that setting. In work co-authored with Daniel Reck, I consider binary choice problems in which preference-irrelevant “frames” affect the behavior of decision-makers and develop an empirical framework for identifying decision-makers’ ordinal preferences given limited data. We show that preference identification hinges upon understanding the empirical relationship between decision-makers’ preferences and their consistency. By recasting the behavioral preference recovery problem in these terms, familiar insights from the program evaluation literature can be fruitfully adapted to this new setting. We illustrate our proposed techniques with data from a range of recent empirical studies.
Acknowledgements

I am grateful to my advisors, Hank Farber and David Lee, for providing encouragement, guidance, and support. Their insightfulness and good humor made my time in the Industrial Relations Section one of the best parts of graduate school. I also thank Alex Mas for a constant supply of good advice over the years.

At Princeton, I was lucky to have a number of amazing teachers. Among them, I am especially grateful to Anne Case, Angus Deaton, Bo Honore, and Harvey Rosen.

My wife Debi moved with me to New Jersey and spent a year waking up at four in the morning to sell coffee drinks at Starbucks. She deserves accolades.

Finally, I couldn’t imagine a better graduate student cohort to be a part of. I owe a debt of thanks to the many who helped keep life interesting, especially Ben Brooks, Karyne Charbonneau, Eliav Danziger, Nicholas Lawson, Dean Spears, and most of all to my first-year study group, Peter Buisseret, Nikolaj Harmon, Tatiana Homonoff, and Rohit Lamba. I wouldn’t be here without them. (Really.) And of course, this dissertation would not exist were it not for my brilliant coauthors, Tatiana Homonoff and Daniel Reck; working with them has set an impossibly high bar for future collaboration.
Contents

Optimal Tax Salience 1

Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity 28

Preference Identification Under Inconsistent Choice 90

Bibliography 170
Chapter 1

Optimal Tax Salience

Abstract

Recent empirical work finds that consumers under-account for commodity taxes when the after-tax price is not prominent. I investigate how policymakers may utilize such “low-salience” taxes to promote welfare. The optimal combination of high- and low-salience taxes balances two competing effects: low-salience taxes dampen distortionary substitution but cause consumers to misallocate their budgets. Using a stylized model, I show the availability of taxes with differing salience provides an extra degree of freedom that can be used to implement the first-best welfare outcome. I characterize the optimal policy and derive a formula for incremental adjustments when the first-best is unattainable.

Introduction

Optimal commodity taxation is a classic subject in the public finance literature. Most research in the area studies how governments should levy taxes across distinct goods to promote social welfare when lump-sum taxes are unavailable. In contrast, questions relating to tax design have not received the same degree of theoretical attention.¹

Recent empirical work suggest a need to reconsider this emphasis. A series of findings suggests that the salience of a tax has important effects on consumer behavior: the less prominent the after-tax price of a good, the less consumers respond to changes in the tax on that good.²

¹By “tax design,” I mean policy decisions relating to characteristics of a tax that do not directly enter into consumers’ budget constraints. Two exceptions are Slemrod and Kopczuk (2002) and Krishna and Slemrod (2003).

²Throughout, I use “salience” to refer to the prominence of the taxed good’s tax-inclusive price. For
Such findings suggest an additional margin through which governments can shape the behavioral effects of a tax. Although policymakers typically lack perfect control over a tax’s salience, they frequently face a choice between relying on high- and low-salience ways of raising revenue. For example, policymakers can manipulate the salience of a commodity tax by choosing whether to include the tax in the displayed price of the taxed good or to add it on at the register when the consumer completes her purchase. Because the former is more salient than the latter, the government can alter the tax’s salience by adjusting the degree to which it relies on the two tax designs.³

This paper studies the optimal salience of commodity taxes: how should a benevolent government choose between high- and low-salience taxes on a particular good to raise revenue?⁴ The analysis highlights two distinct mechanisms through which tax salience affects consumers’ well-being. On the one hand, low-salience taxes dampen the type of excess burden traditionally associated with distortionary taxation: because consumers are less prone to substitute away from goods subject to low-salience taxes, such taxes are less distortionary for a given amount of revenue raised. On the other hand, low-salience taxes drive taxpayers to make optimization errors, reducing welfare by causing consumers to

³Policymakers may also manipulate commodity tax salience by adopting tax-inclusive pricing regulations, which require retailers to include the full amount of consumption taxes in the prices displayed to consumers. Such regulations are common in Europe but are rare in the United States. Similarly, governments may require tax-inclusive pricing for a particular good. For example, the Federal Trade Commission requires airlines to include taxes and other fees in the initial price displayed to consumers. Policymakers may also shape salience in other contexts: road tolls can be collected manually by cash transfers or automatically through an EZ-Pass system (Finkelstein, 2009); property tax payments may be collected on their own or bundled into a monthly mortgage payment to an escrow account (Hayashi, 2014; Cabral and Hoxby, 2013); and income tax payments may be collected from employees or automatically withheld (Jones, 2010).

⁴Addressing this question through the lens of economic theory is complicated by the conceptual and methodological challenges that arise in behavioral welfare analysis. To address these issues, I follow Chetty, Looney and Kroft (2009) and adopt a “refinement” approach to welfare analysis, drawing on insights from Bernheim and Rangel (2009). As I elaborate below, this framework allows one to conduct welfare analysis while remaining agnostic about the exact mechanism driving taxpayer under-reaction to low-salience taxes, at least within a broad range of plausible models.
misallocate income among consumption goods. The government’s choice between high- and low-salience taxes trades off between these competing effects.

In the standard model, the presence of an untaxed good causes the optimal policy to diverge from the first-best; commodity taxes generate excess burden by distorting consumers’ decisions between taxed and untaxed goods. In contrast, when the government can control the salience of a tax, that flexibility provides an additional degree of freedom. I show that when the government can utilize two taxes on a single good that differ in their salience, it can employ those taxes in combination to achieve the first-best welfare outcome – even when one of the available goods cannot be taxed. The key insight is that by adjusting the balance between high- and low-salience taxes, the government can maintain a given level of revenue while causing taxpayers to vary their consumption of the taxed and untaxed good; in this way, taxpayers can be induced to choose the same allocation they would choose under a lump-sum tax (even though that allocation is privately sub-optimal given the taxes that are actually in place).

I next turn to characterizing the optimal combination of high- and low-salience taxes. Solving the government’s problem yields an intuitive formula for the optimal policy, which highlights the link between optimal salience and the nature of demand for the good being taxed. Notably, the formula implies that the optimal size of the low-salience tax is always non-zero. Although low-salience taxes drive consumers to make optimization errors, the welfare costs of those errors is second-order for small values of the tax. In contrast, even small values of a low-salience tax may raise substantial revenues, allowing the government to reduce distortionary high-salience taxes while still meeting its budget constraint.

In practice, adopting policies that are designed to exploit people’s biases raises several important concerns. Although many of these, such as political transparency and credibility, are outside the scope of this paper, one that can undermine the results presented here is the possibility that taxpayers will become more attentive to low-salience taxes as the government increases its reliance on them – i.e., as the utility costs of neglecting them grow
larger. Before concluding, I consider an extension of the model to a setting in which the salience of a tax is endogenously related to the tax’s size and derive conditions under which the first-best will be attainable. When the first-best is unattainable, I show how incremental adjustments in the balance between high- and low-salience taxes can still yield efficiency gains.

Despite the ubiquity of policy decisions that affect tax salience, the topic has received little theoretical attention. As Congdon, Kling and Mullainathan (2009) conclude in their review of the behavioral tax literature, “the theoretical literature has yet to yield the type of rules of thumb with respect to optimal tax salience that translate into practical policy recommendations.” The research closest to the current analysis are Chetty, Looney and Kroft (2009), Chetty (2009a), and Reck (2015). Those authors derive formulas for quantifying the excess burden of a tax that is less than fully salient but do not consider the implications of salience for optimal taxation. In addition, this paper is the first to consider the possibility of combining tax instruments that differ in their salience, and it is that possibility which drives the theoretical insights described here.

A number of influential papers have investigated how cognitive biases other than salience affect prescriptions for optimal tax policy (e.g., Lieberman and Zeckhauser, 2004; O’Donoghue and Rabin, 2003). This literature evaluates the optimal level of a tax instrument conditional on taxpayers exhibiting an assumed behavioral bias. I build on this literature by studying a setting in which the government’s choice of tax instrument controls the extent to which taxpayers exhibit the bias in the first place.

The remainder of the paper proceeds as follows. Section 1 develops the model and derives the main results – first graphically and then formally. Section 2 extends the model to account for the possibility that a tax’s salience is endogenously related to its size. Section 3 concludes.
1. Model and Results

Society is composed of a representative taxpayer who divides her income $I$ between goods $x$ and $y$. Production of $x$ is characterized by constant returns to scale technology so that its pre-tax price is fixed at marginal cost $p$. Good $y$ is the numeraire. The taxpayer’s utility depends on consumption of $x$ and $y$:

$$U = U(x, y)$$  \hspace{1cm} (1.1)

$U$ is concave and smooth with respect to each good.

The government’s objective is to maximize the representative taxpayer’s utility while raising revenue $R_0$.

A. First-Best Welfare Outcome

Before turning to tax salience, it is helpful to characterize the first-best welfare outcome – i.e., what the government could achieve with access to a non-distorting tax. To derive this benchmark I will assume for purposes of this section that the government may levy a (fully-salient) lump-sum tax of size $L$.

When facing the lump-sum tax, the taxpayer’s budget constraint is given by

$$px + y = I - L$$  \hspace{1cm} (1.2)

and her consumption satisfies the first-order condition associated with maximizing her util-

---

5Expressing utility as a function of consumption is standard in public finance models but implies that other factors (such as tax salience) do not affect welfare apart from their effect on consumption. For example, an agent would violate the assumption if she preferred facing a register tax to a posted tax on political grounds, perhaps because the amount going to the government is more transparent under the former than the latter. If low-salience taxes generate direct welfare costs for consumers (independent of their effect on consumption) the results presented here will over-state the benefits of low-salience taxes by neglecting such costs. However, as Chetty, Looney and Kroft (2007) demonstrate, even relatively small cognitive costs generate substantial under-reaction to a tax; as a result, omitting such costs from the model may not be as misleading as would otherwise be the case. Finally, not all psychic cost models are ruled out: suppose that accounting for a low-salience tax requires a consumer to suffer some cognitive cost, but because of that cost, the consumer rationally chooses to ignore the tax. This agent’s utility function can be described by (1.1) because given her decision-making strategy, she does not suffer any direct utility cost when confronted with the tax.
ity subject to this constraint:

\[ U_x(x, y) = pU_y(x, y) \]  

(1.3)

Because the revenue collected by a lump-sum tax of size \( L \) is simply \( L \), the government’s revenue constraint is satisfied if and only if

\[ L = R_0 \]  

(1.4)

Equations (1.2)-(1.4) pin down consumption under a lump-sum tax and hence characterize the first-best welfare outcome.

**B. Tax Salience**

Having characterized the first-best, I assume now that the government lacks access to a lump-sum tax and can only raise revenue through commodity taxes on \( x \). Good \( y \) (the numeraire) is left untaxed. The government has at its disposal two tax designs that it can levy on purchases of \( x \): a high-salience tax \( t_h \) and a low-salience tax \( t_l \). Both \( t_h \) and \( t_l \) are unit taxes. The taxpayer’s budget constraint takes the form:

\[ y + (p + t_h + t_l) x = I \]  

(1.5)

**Taxpayer Behavior**

Taking income as fixed, demand for \( x \) and \( y \) can be written as a function of the two taxes and the pre-tax price of \( x \): \( x = x(p, t_h, t_l) \) and \( y = y(p, t_h, t_l) \). To capture the empirical findings described in the introduction, I assume that the extent to which a tax affects consumer demand depends on the tax’s salience. As in Chetty, Looney and Kroft (2009), I adopt a functional definition of tax salience: for \( i \in \{h, l\} \), the salience of a tax \( (\theta_i) \) measures how taxpayers adjust their demand for the taxed good in response to a change in the tax \( (t_i) \)
relative to a change in the taxed good’s pre-tax price \((p)\):\(^6\)

\[
\theta_h = \frac{\partial x/\partial p}{\partial x/\partial t_h} \quad \theta_l = \frac{\partial x/\partial p}{\partial x/\partial t_l} \tag{1.6}
\]

To illustrate the notation, a tax that appeared as part of the taxed good’s posted price (e.g. an excise tax) would be fully-salient (i.e., \(\theta = 1\)). In contrast, a tax to which consumers were entirely unresponsive would have \(\theta = 0\). I assume that the two tax designs available to the government have differing (but individually-fixed) degrees of salience; the taxpayer is more responsive to changes in the high-salience tax than to changes in the low-salience tax:

\[
0 \leq \theta_l < \theta_h \leq 1 \tag{1.7}
\]

Whether Equation (1.7) is satisfied in a particular context is an empirical question. One common situation in which (1.7) will be satisfied is when the government has access to one commodity tax instr that is less than fully salient \((\theta_l < 1)\), such as a sales tax, and another that directly affects the posted price of the taxed good \((\theta_h = 1)\), such as an excise tax. Equation (1.7) also imposes that the salience of the two tax instruments are between 0 and 1.

Finally, I assume throughout that \(x + (t_l + t_l) \frac{\partial x}{\partial t_i} \geq 0\) for \(i \in \{h,l\}\).\(^7\)

**Behavioral Welfare Framework**

Under the standard neoclassical model, demand for \(x\) and \(y\) correspond to the solution to the consumer’s welfare maximization problem: \(\text{MAX}_{x,y} \; U(x,y) \; s.t. \; y + (p + t_h + t_l) x = I\). Let \(x^*(p,t_h,t_l)\) and \(y^*(p,t_h,t_l)\) denote the quantities of \(x\) and \(y\) that solve this problem, i.e.,

\(^6\)This functional definition of salience corresponds to the common understanding of salience as “prominence” if the pre-tax price of a good is prominent to consumers and consumers fully account for the pre-tax price when making purchasing decisions. Equation (1.6) assumes a constant degree of under-reaction by consumers when a tax is less than fully-salient. I relax this assumption in Section 2.

\(^7\)The case in which \(x + (t_h + t_l) \frac{\partial x}{\partial t_l} \leq 0\) is uninteresting because it implies that \(\frac{\partial x}{\partial t_l} = x + (t_h + t_l) \frac{\partial x}{\partial t_l} \leq 0\) so that the government could raise additional revenue solely by reducing one of its tax instruments.
that maximize the taxpayer’s utility subject to the budget constraint. It is straightforward to show that \( x^* \) and \( y^* \) depend only on the total after-tax price of the taxed good, \( x^*(p, t_h, t_l) = x^*(p + t_h + t_p) \) and \( y^*(p, t_h, t_l) = y^*(p + t_h + t_p) \). Yet as discussed in Chetty, Looney and Kroft (2009), this result is inconsistent with the empirical evidence that consumer behavior depends in part on tax salience (rather than the size of the tax alone). Consequently, the neoclassical revealed preferences approach to welfare analysis – which assumes rational decision-making by consumers – is arguably inappropriate for analyzing policy decisions about tax salience.

Instead, I follow Chetty, Looney and Kroft (2009) by utilizing what Bernheim and Rangel (2009) refer to as a “refinement.” Rather than assume that every decision the taxpayer makes reflects her true preferences, I assume the taxpayer behaves optimally when tax-inclusive prices are fully salient (e.g., when all taxes are included in the posted price) \( \text{Equation (1.8)} \)

\[
x(p + t_h + t_l, 0, 0) = x^*(p + t_h + t_l)
\]

**Government’s Problem**

As above, I consider the problem faced by a government seeking to maximize consumer welfare subject to a revenue constraint \( R_0 \). It will be convenient to express consumer welfare as a function of the government’s choice of taxes:

\[
V(t_h, t_l) = U(x(t_h, t_l), y(t_h, t_l))
\]

Total government revenue \( R \) is also a function of the taxes: \( R(t_h, t_l) = (t_h + t_l) x(p, t_h, t_l) \). The government’s revenue constraint is therefore given by

\[\text{Equation (1.8)}\]

\[\text{Equation (1.9)}\]

---

\(^8\)That is, if consumers behave optimally, Equations (1.5) and (1.1) are inconsistent with (1.6) and (1.7).

\(^9\)Equation (1.8) is a weaker version of the typical rationality assumption underlying the revealed preference approach to welfare analysis. Rather than assume that all of a decision-maker’s choices reflect her true preferences, this approach imposes rationality only for the subset of choices made when taxes are fully salient. Because there are good reasons to be skeptical about the quality of choices made when taxes are less than fully-salient, this approach privileges the preferences revealed when those conditions are not present.
The government’s problem is to choose the combination of \( t_h \) and \( t_l \) that solves

\[
(t_h + t_l) x(p, t_h, t_l) = R_0
\]  

(1.10)

\[ \text{MAX}_{t_h, t_l} V(t_h, t_l) \ s.t. \ (t_h + t_l) x(p, t_h, t_l) = R_0 \]  

(1.11)

C. Graphical Illustration

This section provides graphical intuition for the main result of the paper: that having access to two taxes with differing salience provides the government an extra degree of freedom that it can use to implement the first-best welfare outcome. The next section provides a formal proof.

Consider a stylized example, depicted in Figure 1. Suppose the government is choosing between a fully-salient tax (\( \theta_h = 1 \)) and a tax to which consumers are entirely unresponsive (\( \theta_l = 0 \)). The consumer’s pre-tax budget constraint is given by the line AB, and pre-tax consumption (\( E_0 \)) is characterized by the tangency of the consumer’s indifference curve with AB. Because any feasible choice of taxes must raise revenue \( R_0 \), the taxpayer’s consumption under any feasible tax combination will lie somewhere on the line CD, which is simply line AB shifted downwards by the vertical distance \( R_0 \).

To identify consumption under the first-best, consider a lump-sum tax of size \( R_0 \). Because lump-sum taxes do not affect relative prices, the budget constraint induced by the lump-sum tax is also given by line CD. Consumption under the first-best allocation (\( E_{LST} \)) is determined by the point at which the consumer’s indifference curve is tangent to line CD.

Now, suppose the government relies solely on \( t_h \) for raising revenue. In that case, the consumer’s budget constraint would pivot to line AF. Consumption under this tax (\( E_h \)) is the tangency point between line AF and the consumer’s indifference curve. Note that the high salience tax generates excess burden by driving consumers to substitute away from the taxed good over and above the income effect of the tax, \( x_h < x_{LST} \).
In contrast, if the government were to rely solely on $t_l$ for raising revenue, demand for the taxed good would not change from its pre-tax level, inducing consumption at $E_l$. Although consumers do not substitute away from the taxed good, the tax still generates an excess burden because consumers fail to adjust their consumption to account for the tax’s income effect, $x_l > x_{LST}$.

Intuitively, because all feasible tax combinations induce consumption along the line CD, and because consumption under the lump-sum tax lies between the consumption induced by the high- and low-salience taxes (when either is imposed alone), the optimal policy is somewhere between full reliance on either $t_h$ or $t_l$. That is, by shifting the balance between the high- and low-salience tax, the government can move consumption along CD until it reaches $E_{LST}$. In particular, suppose the government imposes the high salience tax at a level – call it $t^*_h$ – that pivots the consumer’s budget constraint to line AG, thereby inducing consumption at $E_s$, where demand for the taxed good is equal to demand for the taxed good under the first-best. Since $E_s$ lies above line CD, this tax, on its own, fails to meet the government’s revenue constraint. However, the government can combine $t^*_h$ with a low-salience tax ($t^*_l$) to make up the additional revenue. And because $\theta_l = 0$, imposing the low-salience tax does change the amount of $x$ demanded by the taxpayer; it simply shifts consumption downwards. As a result, the government can combine $t^*_h$ and $t^*_l$ to induce $E_{LST}$ – the same consumption that would be induced under a lump-sum tax.\(^{10}\)

This simple example illustrates how the availability of multiple tax instruments that differ in their salience provides policymakers an additional degree of freedom with which to shift consumer demand while maintaining a desired level of revenue. The following sections formalize this intuition and explore the conditions that must be met for the first-best to be attainable.

\(^{10}\)Note that although $E_{LST}$ is the socially-optimal allocation, it is privately sub-optimal under $t^*_h$ and $t^*_l$. That is, any individual taxpayer facing $t^*_h$ and $t^*_l$ would be (privately) better off consuming an allocation with less $x$ and more $y$ than at $E_{LST}$.\)
D. Welfare Under Optimal Tax Salience

To derive the optimal policy, I begin with an arbitrary (feasible) combination of high- and low-salience taxes and consider the welfare consequences of (feasible) adjustments to the initial combination. The optimal policy corresponds to the combination of taxes for which no feasible adjustment would generate a positive welfare benefit.

Consider some combination of taxes \((t_h, t_l)\) that satisfies the government’s revenue constraint, \((t_h + t_l) x(p, t_h, t_l) = R_0\). The government may adjust \(t_h\) and \(t_l\), but in order for the combination to be feasible, it must adjust the taxes in such a way that the revenue constraint continues to hold. The feasible combinations of taxes can be found by totally differentiating the revenue constraint:

\[
\frac{\partial}{\partial t_h} \left( x + (t_h + t_l) \frac{\partial x}{\partial t_h} \right) d t_h + \frac{\partial}{\partial t_l} \left( x + (t_h + t_l) \frac{\partial x}{\partial t_l} \right) d t_l = 0.
\]

With (1.6), this yields the change in the high-salience tax associated with a small increase in the low-salience tax such that the overall policy change is revenue-neutral.

\[
\frac{\partial t_h}{\partial t_l} \bigg|_{R_0} = -\frac{\theta_h \frac{\partial x}{\partial p} (t_h + t_l)}{\theta_l \frac{\partial x}{\partial p} (t_h + t_l) + x},
\]

In words, a $1 increase in the low-salience tax accommodates a revenue-neutral reduction in the high-salience tax of \(\frac{\partial t_h}{\partial t_l} \bigg|_{R_0}\) dollars. Note that (1.7) implies \(\frac{\partial t_h}{\partial t_l} \bigg|_{R_0} < -1\).\(^\text{11}\) Intuitively, because taxpayers adjust their demand more in response to changes in the high-salience tax, a $1 increase in the low-salience tax accommodates a revenue-neutral reduction in the high-salience tax of more than $1.

Define a revenue-neutral shift towards the low-salience tax as a marginal increase in \(t_l\) along with the corresponding reduction in \(t_h\) needed to maintain revenue neutrality. The following result describes the welfare effect of this policy change.

\(^{11}\)In particular, this follows from \(\theta_h > \theta_l\) along with the maintained assumption that \(x + (t_h + t_l) \frac{\partial x}{\partial t_i} > 0\) for \(i \in \{h, l\}\).
Lemma 1  The welfare effect of a revenue-neutral shift towards the low-salience tax, $dV \bigg|_{t_l}^{\frac{dt}{R_0}}$, is given by

$$
\frac{dV}{dt} \bigg|_{R_0} = \left[ \left( -x + \frac{\partial t_l}{\partial t_l} \bigg|_{R_0} \right) + \left( \frac{\theta_l + \partial t_h}{\partial t_l} \bigg|_{R_0} \right) \theta_h \right] \frac{\partial x}{\partial t_l} + \left( \frac{U_x}{U_y} - (p + t_h + t_l) \right) \frac{\partial y}{\partial t_l}
$$

Proof  Totally differentiating the consumer’s budget constraint (1.5) yields

$$
\frac{\partial y}{\partial t_i} = -x - (p + t_h + t_l) \frac{\partial x}{\partial t_i} \quad \text{for } i \in \{h, l\}
$$

(1.13)

Totally differentiating (1.9) yields:

$$
\frac{dV}{dt} \bigg|_{R_0} = U_x(x,y) \frac{\partial x}{\partial t_l} + U_x(x,y) \frac{\partial t_h}{\partial t_l} + U_y(x,y) \frac{\partial y}{\partial t_l} + U_y(x,y) \frac{\partial y}{\partial t_h} \frac{\partial t_h}{\partial t_l} \bigg|_{R_0}
$$

(1.14)

Substituting (1.6) and (1.13) into (1.14) yields the result.

The expression in Lemma 1 is the sum of two intuitive components. Term 1 captures the welfare effect from the shift on the taxpayer’s purchasing power. The net change in taxes on $x$ is given by $\left( \frac{d(t_h + t_l)}{dt_l} \right) \bigg|_{R_0} = 1 + \frac{\partial t_h}{\partial t_l} \bigg|_{R_0}$; scaling that price change by the taxpayer’s consumption of $x$ yields the change in purchasing power. Term 1 is positive because, as explained above, $\theta_l < \theta_h$ guarantees $\frac{\partial t_h}{\partial t_l} \bigg|_{R_0} < -1$ – in words, a revenue-neutral shift towards the low-salience tax accommodates a net reduction in taxes on $x$ and a corresponding increase in the taxpayer’s purchasing power.

The second piece of Lemma 1 (the product of Terms 2, 3, and 4) captures the welfare loss from optimization errors. To interpret Term 2, it is helpful to first define the price-equivalent tax, $p_\theta$, to be the magnitude of the pre-tax price change that would induce the same change in demand for $x$ as imposing $t_l$ and $t_h$, $p_\theta = \theta_l + \theta_h t_l$. Loosely speaking, this quantity reflects the after-tax price of $x$ that is perceived by consumers. Term 2 is
therefore equal to the change in the price-equivalent tax induced by the shift, \( \frac{d\theta}{dt} \bigg|_{R_0} \), i.e., the change in the price of \( x \) as perceived by taxpayers. Term 3 maps the price-equivalent tax change into behavior; the product of Terms 2 and 3 reflects the increase in consumption of \( x \) induced by the policy shift. Finally, Term 4 maps the change in consumption into welfare. Note that the first-order condition associated with the taxpayer’s welfare maximization problem implies that Term 4 equals 0; thus if the taxpayer’s behavior were optimal, a marginal increase in consumption of \( x \) would have no effect on welfare. In contrast, when taxpayers misperceive low-salience taxes, consumption of the taxed good is sub-optimally high. In particular, Term 4 is negative because the marginal utility of expenditures on \( y \) exceeds the marginal utility of expenditures on \( x \). Taken as a whole, the product of Terms 2, 3, and 4 capture the fact that shifting towards the low-salience tax causes taxpayers to incur welfare losses by over-consuming the taxed good.

Lemma 1 highlights the tension that characterizes the optimal salience problem: increasing the government’s reliance on low-salience taxes reduces the total taxes on \( x \) (raising consumers’ purchasing power), but induces consumers to deviate further from their optimal consumption bundle. Note that were \( \theta_h = \theta_l \) (in violation of Assumption (1.7)), both pieces of Lemma 1 would be equal to zero; intuitively, when both of the available tax instruments have the same salience, shifting between them does not affect welfare.

Under the optimal policy, no (feasible) shift in taxes generates an improvement in welfare, \( \frac{dV}{dt} \bigg|_{R_0} = 0 \). This condition allows us to characterize the first-order condition to the government’s problem.

**Lemma 2** The optimal combination of high- and low-salience taxes induces the taxpayer to consume values of \( x \) and \( y \) that satisfy

\[
U_x(x,y) - pU_y(x,y) = 0
\]

**Proof** At the optimum, no feasible shift in taxes generates an improvement in welfare,
\[
\frac{dV}{dt}_{|_{R_0}} = 0.
\] Setting the expression in Lemma 1 equal to zero and substituting in the expression for \( \frac{\partial h}{\partial t}_{|_{R_0}} \) from (1.12) yields:

\[
-x \left( 1 - \frac{\theta_l \frac{\partial x}{\partial p}(t_h + t_l) + x}{\theta_h \frac{\partial x}{\partial p}(t_h + t_l) + x} \right) + \left( \theta_l - \theta_h \frac{\partial x}{\partial p}(t_h + t_l) + x \right) \frac{\partial x}{\partial p} \left( \frac{U_x}{U_y} - (p + t_h + t_l) \right) = 0
\]

or, after simplifying:

\[
(\theta_h - \theta_l) \left( \frac{U_x}{U_y} - p \right) = 0 \tag{1.15}
\]

Assumption (1.7) guarantees \( \theta_h - \theta_l \neq 0 \); dividing (1.15) by that quantity yields the result.\(^{12}\)

Comparing the conditions that characterize consumption under the lump-sum tax with the conditions that characterize consumption under the optimal combination of high- and low-salience taxes yields the main result of this section.

**Proposition 1** The optimal combination of high- and low-salience taxes achieves the first-best welfare outcome.

**Proof** Consumption under the lump-sum tax is determined by (1.2), (1.3), and (1.4); consumption under the optimal combination of high- and low-salience taxes is determined by (1.5), (1.10), and Lemma 2. Equations (1.2) and (1.4) imply \( px + y = I - R_0 \), which is also implied by (1.5) and (1.10). In addition, (1.3) is identical to Lemma 2. Hence, consumption under the lump-sum tax is equal to consumption under the optimal combination of \( t_h \) and \( t_l \). Because utility depends only on consumption (1.1), welfare under the two policies is the same as well.

Proposition 1 demonstrates that control over tax salience – in the sense of having available two taxes of differing salience – provides the government an extra degree of freedom with which to implement the first-best, even when one of the available goods cannot be taxed.

\(^{12}\) A feasible allocation that satisfies Lemma 2 is guaranteed to be the optimum because it induces the first-best welfare outcome, as shown in Proposition 1 (below).
As illustrated in Section 1.C, the basic intuition is that the first-best level of demand for the taxed good can be induced by different combinations of the tax instruments, but each combination yields a different amount of revenue. Implementing the first-best thus requires identifying which of these combinations yields sufficient revenue to meet the government’s revenue constraint.\footnote{This basic theoretical insight continues to hold even when a lump-sum tax does not yield the first-best allocation. For example, it is straightforward to show that the optimal combination of high- and low-salience taxes can also achieve the first-best in settings where the taxed good generates a consumption externality.}

**E. Characterizing the Optimal Salience of a Tax**

This section investigates what combination of high- and low-salience taxes are required to implement the optimal policy. I first derive a formula for the optimal policy by generalizing the graphical approach described in Section 1.C. I then relate the optimal degree of salience to observable elasticities, which sheds light on the economic forces that shape the optimal policy.

To combine high- and low-salience taxes in a manner that implements the first-best, consider the following approach. First, set $t_h = \overline{t}_h$, where $\overline{t}_h$ is defined as the level of the high-salience tax that, when employed without any other taxes, induces the first-best quantity of consumption of $x$. In Figure 1, $\overline{t}_h$ corresponds to $t^*_h$.\footnote{In Figure 1, $\overline{t}_h$ was not only the value of the high salience tax that induced the first-best consumption of $x$, it also happened to be the high salience tax’s optimal value. More generally, when $\theta_l \neq 0$ or when $\theta_h < 1$, $\overline{t}_h$ may diverge from the optimal high salience tax, as described below.} Writing demand for $x$ as a function of $t_h, t_l$, and after-tax income, the first-best can be written as the value of $x$ induced by a lump-sum tax of size $R_0$, $x_{LST} \equiv x(0,0,I - R_0)$. Consequently, $\overline{t}_h$ is implicitly defined by $x(\overline{t}_h,0,I) \equiv x(0,0,I - R_0)$.

Although $\overline{t}_h$ induces the first-best level of consumption of $x$, the revenue raised by $\overline{t}_h$ will not in general be equal to $R_0$. However, the difference in salience between the available tax instruments permit the government to adjust the balance between $t_h$ and $t_l$ in ways that increase revenue but do not affect consumption of $x$. That is, by combining increases in $t_l$
with reductions in \( t_h \), the government can increase revenue without causing individuals to substitute away from the taxed good. To see this, note that totally differentiating \( x \) (while holding after-tax income fixed) yields:

\[
dx = \frac{\partial x}{\partial t_l} \partial t_l + \frac{\partial x}{\partial t_h} \partial t_h = \theta_l \frac{\partial x}{\partial p} \partial t_l + \theta_h \frac{\partial x}{\partial p} \partial t_h
\]

It follows that movements along the line \( \frac{\partial t_l}{\partial t_h} = -\frac{\theta_h}{\theta_l} \) do not affect the taxpayer’s demand for \( x \). On the other hand, movements along this line do affect the amount of revenue that is raised:

\[
dR = (t_l + t_h) \frac{\partial x}{\partial t_h} + x_{LST} \left( 1 + \frac{\partial x}{\partial t_h} \right).
\]

Because \( \frac{\partial x}{\partial t_h} = 0 \), we have:

\[
dR = -x_{LST} \left( \frac{\theta_h - \theta_l}{\theta_l} \right)
\]

Thus for each $1 reduction in \( t_h \), \( t_l \) may be increased by \( \frac{\theta_h}{\theta_l} \) dollars without causing consumption of \( x \) to depart from \( x_{LST} \). At the same time, this policy change raises revenue in the amount of \( x_{LST} \left( \frac{\theta_h - \theta_l}{\theta_l} \right) \) dollars.

Suppose the government initially sets \( (t_h, t_l) = (\overline{t_h}, 0) \) and subsequently reduces \( t_h \) by \( \delta \) dollars while increasing \( t_l \) by \( \frac{\theta_h}{\theta_l} \delta \) dollars. The net result of this policy is that consumers choose \( x = x_{LST} \), and the total amount of revenue raised is \( R = \overline{t_h} x_{LST} + \delta x_{LST} \left( \frac{\theta_h - \theta_l}{\theta_l} \right) \). Setting \( R = R_0 \) allows us to solve for the value of \( \delta \) that raises the required amount of revenue:

\[
\delta^* = (\tau - \overline{t_h}) \left( \frac{\theta_l}{\theta_h - \theta_l} \right)
\]

where \( \tau \equiv \frac{R_0}{x_{LST}} \).

Using (1.16), we can now solve for \( t_h^* = \overline{t_h} - \delta^* \) and \( t_l^* = \frac{\theta_h}{\theta_l} \delta^* \), the values of the taxes that induce first-best consumption while satisfying the revenue constraint.

\[
t_h^* \equiv \overline{t_h} - \delta^* = \left( \frac{\theta_h \overline{t_h} - \theta_l \tau}{\theta_h - \theta_l} \right)
\]

\[
t_l^* \equiv \frac{\theta_h}{\theta_l} \delta^* = (\tau - \overline{t_h}) \left( \frac{\theta_l}{\theta_h - \theta_l} \right)
\]
Equations (1.17) and (1.18) allow one to implement the first-best solution given knowledge of $x_{LST}$ and $\pi_t$. To implement the optimum when these quantities are not known, and to better understand the mechanisms at work, it is helpful to express $t^*_h$ and $t^*_l$ as functions of more familiar quantities. Let $\eta_{x,l} = \frac{\partial x}{\partial I} I$ denote the income-elasticity of $x$, $\omega_x = \frac{p_x}{I}$ the budget share of expenditures on $x$, and $\epsilon_{x,p} = -\frac{\partial x}{\partial p} p$ the own-price elasticity of $x$ (defined to be positive), where each quantity is evaluated at the no-tax baseline.

Define

$$\theta^* = \frac{\eta_{x,l} \omega_x}{\epsilon_{x,p}}$$

The numerator of $\theta^*$ represents the income effect associated with a price increase on $x$ – the elasticity corresponding to the slope of the Engel curve through the no-tax optimum. The denominator of $\theta^*$ represents the combined income and substitution effects, $\epsilon_{x,p} = \tilde{\epsilon}_{x,p} + \omega_x \eta_{x,l}$, where $\tilde{\epsilon}_{x,p}$ denotes the compensated (Hicksian) own-price elasticity of demand. By scaling the income effect by the combined income and substitution effects, a tax with salience $\theta^*$ induces taxpayers to adjust their consumption of $x$ as if there were no substitution effect, but to still account for the tax’s income effect (as they would under a lump-sum tax).

The following proposition shows that $\theta^*$ describes the optimal salience for taxes on $x$, in the following sense. When the government has available to it a tax instrument (either $t_h$ or $t_l$) with salience exactly equal to $\theta^*$, the optimal policy is to rely on that instrument exclusively. When no such tax is available, the government may replicate the welfare effects of a tax with optimal salience by combining the tax instruments that are available so that the weighted average replicates a single tax with salience $\theta^*$.

**Proposition 2** Let $\rho$ denote the fraction of taxes on $x$ that are low-salience: $\rho = \frac{t_l}{t_h + t_l}$. Let $\theta^* = \frac{\eta_{x,l} \omega_h}{\epsilon_{x,p}}$, where each quantity is evaluated at the no-tax baseline. Then the optimal combination of high- and low-salience taxes is given by the value of $\rho$ that solves

$$\rho \theta_l + (1 - \rho) \theta_h = \theta^*$$

17
**Proof** By the definition of $\bar{t}_h$, we have that $x(0,\bar{t}_h,I) \equiv x(0,0,I - R_0)$. Subtracting $x(0,0,I)$ from both sides and applying a first-order Taylor approximation yields:

$$\bar{t}_h \theta_h \frac{\partial x}{\partial p} \approx -R_0 \frac{\partial x}{\partial I} \quad (1.19)$$

Using the definitions of $\eta_{x,I}$, $\omega_x$, $\varepsilon_{x,p}$, and $\tau$, it is straightforward to re-write (1.19) as

$$\bar{t}_h \approx \left( \frac{\eta_{x,I} \omega_x}{\varepsilon_{x,p}} \right) \frac{\tau}{\theta_h} \quad (1.20)$$

Substituting (1.20) into (1.17) and (1.18) yields an expression for the optimal taxes in terms of $\theta^*$.

$$t^*_h \approx \frac{\tau}{\theta_h - \theta_l} (\theta^* - \theta_l) \quad (1.21)$$

$$t^*_l \approx \frac{\tau}{\theta_h - \theta_l} (\theta_h - \theta^*) \quad (1.22)$$

Finally, noting that $t^*_h + t^*_l = \tau$, we can re-write (1.22) in terms of $\rho$ to obtain the result.

Proposition 2 yields a number of important insights. First, as discussed above, $\theta^*$ represents the ratio of the tax’s income effect to its combined income and substitution effects; a tax with salience $\theta^*$ therefore induces taxpayers to alter their consumption as if there were no substitution effect but only an income effect associated with the tax. In other words, introducing a new tax with salience $\theta^*$ to raise a marginal amount of revenue induces taxpayers to reduce their demand for $x$ by following the Engel curve through the no-tax optimum – thereby replicating (locally) the behavioral effects of a lump-sum tax.\(^{15}\)

\(^{15}\)To see this formally, write $\theta^* = -x \frac{\partial \varepsilon_{x,\rho}}{\partial \eta_{x,p}}$ so that a tax $t$ with salience $\theta^*$ induces a change in demand of $\frac{\partial x}{\partial t} = \theta^* \frac{\partial x}{\partial p} = -x \frac{\partial x}{\partial t}$. The effect on $x$ of raising a marginal amount of revenue using $t$ is given by $\frac{\partial x}{\partial R} = \frac{\partial x}{\partial R} = \frac{-x \partial x}{x - \rho \frac{\partial x}{\partial t}} = \frac{-\frac{\partial x}{\partial t}}{\theta^*} \frac{\partial x}{\partial t}$ at the no-tax baseline $t = 0$. Additionally, the effect on $x$ of raising a marginal amount of revenue using lump-sum tax $L$ is also given by $\frac{\partial x}{\partial R} = \frac{\partial x}{\partial L} = -\frac{\partial x}{\partial L}$. 

18
Second, for normal goods, $\theta^* \in [0, 1)$. This implies that even when the government has access to fully salient tax ($\theta_h = 1$), the optimal policy is not to rely on it exclusively (i.e., $\rho > 0$) despite the fact that doing so would eliminate any mistakes by taxpayers. This result can be readily understood as an application of the Theory of the Second Best (Lipsey and Lancaster, 1956-67). That is, the government’s need to raise revenue through a commodity tax creates a distortion that pushes social welfare away from the first-best welfare outcome. Consequently, by introducing a new distortion – taxpayer deviations from optimal decision-making – policymakers can increase social welfare. More generally, Proposition 2 implies that the optimal policy involves utilizing both of the available tax instruments (outside of the knife-edge case in which the salience of one of the available taxes is exactly equal to $\theta^*$).

Third, Proposition 2 shows that the optimal combination of high- and low-salience instruments depends upon the nature of demand for the good being taxed. In particular, $\theta^*$ is declining in $\tilde{\varepsilon}$. Intuitively, the excess burden associated with a tax depends on the compensated elasticity of the taxed good (Auerbach, 1985); the greater is $\tilde{\varepsilon}$, the larger the welfare gains from reducing the consumer substitution that is typically associated with commodity taxes in the presence of an untaxed good. Additionally, $\theta^*$ is increasing in $\eta_{x,I}$ – the income elasticity associated with the taxed good. This is because the welfare cost of the budgeting mistake is larger for goods with higher income elasticities – neglecting the reduction in purchasing power caused by the tax leaves a taxpayer further from the amount of $x$ she would consume at her (private) optimum. Note that $\theta^* = 0$ if and only if demand for the taxed good is entirely insensitive to income.\footnote{Mechanically, this follows from the fact that $\tilde{\varepsilon}_{x,p} > 0$ when consumers behave optimally, and (1.8) guarantees that consumers behave optimally at the no-tax baseline, where $\tilde{\varepsilon}$ is evaluated.}

\footnote{To understand the intuition, consider a tax to which consumers are entirely unresponsive ($\theta = 0$). Let $(x_0, y_0)$ represent the taxpayer’s initial consumption of $x$ and $y$ at tax $t^0$. Suppose the government raises the tax to $t^1 = t^0 + \alpha$. Because $\theta = 0$, consumers buy the same amount of $x$ as before the tax increase, leaving them with $\alpha x_0$ less income to spend on other goods. When $\eta_{x,I} = 0$, this response exactly matches how a fully-optimizing agent would respond to the tax. Because the optimal choice of $x$ does not depend on income, the consumer has nothing to gain by reconsidering her consumption of $x$ after a decline in income. In contrast, when $\eta_{x,I} > 0$, the consumer who fails to adjust her consumption of $x$ in response to a tax increase is worse off for failing to do so. See Chetty, Looney and Kroft (2009) for a closely related discussion.}
Finally, Proposition 3 highlights the conditions under which subsidies will be required to implement the first-best. In particular, the optimal value of both tax instruments will be non-negative if and only if \( \theta^* \in [\theta_l, \theta_h] \).\(^{18}\) When \( \theta^* < \theta_l \), Equations (1.21) and (1.22) show that implementing the first-best requires utilizing a low-salience tax in conjunction with high-salience subsidy.\(^{19}\) When subsidies are unavailable and \( \theta^* \notin [\theta_l, \theta_h] \), it is straightforward to show that the optimal policy takes the form of a corner solution, in which the government relies solely on the tax that has salience closest to \( \theta^* \).

2. Optimal Policy When Salience is Endogenous

So far I have assumed that the degree of salience associated with the available tax instruments is fixed and exogenous to the model. In practice, it may be that a tax’s salience depends in part on its size. For example, some bounded rationality models of decision-making imply that consumers will pay more attention to larger taxes because the utility costs of neglecting a large tax are greater than those from neglecting a small tax (Chetty, Looney and Kroft, 2007; Reck, 2015).\(^{20}\) When taxpayers behave in this way, the salience of the tax will be increasing in the tax’s size and the results from previous sections may not hold.

\(^{18}\)Mechanically, this follows from (1.21) and (1.22). An immediate implication is that the first-best is always attainable without subsidies when the government has available to it taxes with salience \( \theta_l = 0 \) and \( \theta_h = 1 \).

\(^{19}\)This assumes the salience of a tax instrument is identical to the salience of a similarly-designed subsidy. In at least some contexts, tax subsidies may have higher salience than similarly-designed taxes (Feldman and Ruffle, 2015). Additionally, when subsidies are required to reach the first-best, Equation (1.21) highlights the factors that shape how large the subsidy must be. First, when revenue requirements from taxes on \( x \) \((R_0)\) are large and the amount of \( x \) consumed under the first-best policy \((x_{LST})\) is small, \( \tau \) will be large and hence the required subsidy will tend to be large as well. Second, when the available tax instruments have similar salience, i.e. \( \theta_h \approx \theta_l \), the required subsidy will be quite large. In the extreme case in which \( \theta_h = \theta_l \), the required subsidy would be infinitely large and the result would not hold.

\(^{20}\)On the other hand, researchers have documented behavioral biases even in decision-making contexts where the stakes are large, such as retirement savings decisions (Beshears et al., 2009), high-interest borrowing (Bertrand and Morse, 2011), labor supply decisions by earned income tax credit filers (Chetty and Saez, 2013), and property tax assessment appeals (Hayashi, 2014). In the commodity tax context, Feldman and Ruffle (2015) do not find any evidence that salience effects diminish when moving from low- to high-priced products, even though a consumer’s failure to consider an ad-valorem tax has a larger utility cost in the latter case than in the former.
Feasibility of the First-Best

This section derives conditions for whether the first-best is feasible in settings where tax salience is endogenously related to the size of the tax. Suppose the government has two tax instruments available to it, a high-salience tax with \( \theta_h \) fixed at 1 (such as an excise tax) and a low-salience tax for which the salience depends positively upon the tax’s size, \( \theta_l = \theta_l(t_l) \), \( \frac{\partial \theta_l}{\partial t_l} > 0 \). We can begin as before by setting \( t_h \) at the level necessary to induce consumers to consume \( x \) at the first-best quantity, \( t_h = \bar{t}_h \). As before, consider a reduction in \( t_h \) along with an increase in \( t_l \) so that the net effect is for taxpayers to continue consuming \( x \) at \( x_{LST} \).

Totally differentiating demand for \( x \) yields \( \frac{\partial t_l}{\partial t_h} |_{x_{LST}} = -\theta_l(t_l) \). As a result, the additional revenue generated by an “\( x \)-neutral” increase in \( t_l \) is given by

\[
\frac{dR}{dt_l} |_{x_{LST}} = \frac{\partial (t_l + t_h)}{\partial t_l} |_{x_{LST}} x_{LST} = (1 - \theta_l(t_l)) x_{LST} \tag{1.23}
\]

In order to attain the first-best, the government must be able to increase \( t_l \) (and reduce \( t_h \)) by a sufficient amount to raise \( R_0 \) without altering demand for \( x \). Consequently, the first-best welfare outcome is feasible if and only if there exists a value of the low-salience tax, \( \hat{t}_l \), such that \( \bar{t}_h x_{LST} + \int_0^{\hat{t}_l} \frac{\partial R}{\partial t_l} |_{x_{LST}} \, dt_l \geq R_0 \), or, using Equation (1.23):

\[
\bar{t}_h + \int_0^{\hat{t}_l} (1 - \theta_l(t_l)) \, dt_l \geq \tau \tag{1.24}
\]

where \( \tau = \frac{R_0}{x_{LST}} \).

Finally, as before, \( \bar{t}_h \) can be expressed in terms of familiar quantities by noting that \( x(p, \bar{t}_h, 0, I) \equiv x(p, 0, 0, I - R_0) \). Taking first-order Taylor approximations around \( x(p, 0, 0, I) \)

\footnote{One possibility, inconsistent with Equation (1.1)’s implication that welfare depends solely on consumption, is that taxpayers suffer psychological costs from accounting for low-salience taxes, and that these costs are increasing as attentiveness to the tax increases. In this case, (1.24) being satisfied no longer guarantees that policymakers can reach the first-best; even when the taxes induce consumers to choose the first-best bundle of goods, consumers may be worse-off relative to the first-best because they are suffering the psychological costs associated with paying some attention to the low-salience tax. Note that to the extent that attentiveness to the taxes is (locally) stable, policymakers may still employ the incremental approach described later in this section for adjusting the balance between high- and low-salience taxes.}
implies $\bar{t}_h \vartheta_{\sigma p} \approx -R_0 \vartheta_{\sigma I}$. Writing this result in terms of elasticities yields $\bar{t}_h \approx \tau \theta^*$, where $\theta^*$ is defined as in Proposition 2. Substituting this approximation into (1.24) implies that the optimal combination of high- and low-salience taxes achieves the first-best welfare outcome if and only if

$$\int_0^{\bar{t}_h} (1 - \theta_l(t_l)) \vartheta_{\sigma I} t_l \leq \tau (1 - \theta^*)$$

(1.25)

Thus, when the salience of the available tax instruments is endogenous, determining whether the first-best welfare outcome is feasible depends on three factors. First, because $\tau$ depends positively on $R_0$, the greater the revenue that must be raised from taxes on $x$, the harder it will be to attain the first-best. Second, it will be easier to attain the first-best when the optimal degree of salience for the taxed good is relatively high (large $\theta^*$), e.g., when demand for the taxed good is relatively income elastic. Intuitively, achieving the first-best in such cases requires relying less heavily on the low-salience tax, reducing the likelihood that consumers will become more attentive to it. Finally, whether the first-best can be achieved depends on the relationship between tax size and salience; the slower that $\theta_l(t_l)$ increases when the government increases its reliance on $t_l$, the more likely the first-best will be feasible.

**Local Improvements when the First-Best is not Feasible**

Even when the salience of the available tax instruments increases too fast to achieve the first-best, the results here can still shed light on whether incremental changes in the balance between high- and low-salience taxes is desirable. In particular, suppose that for the current values of $t_h$ and $t_l$, $\theta_h$ and $\theta_l$ are such that $\rho \theta_h(t_h) + (1 - \rho) \theta_l(t_l) > \theta^*$, where $\rho$ and $\theta^*$ are defined as in Proposition 2. In such cases, it is straightforward to show that the welfare effect of an incremental revenue-neutral shift towards the low-salience tax will be positive, and vice-versa when $\rho \theta_h(t_h) + (1 - \rho) \theta_l(t_l) < \theta^*$. This claim is formalized in the Appendix.
Because of this, computing $\theta^*$ and identifying $\theta_l$ and $\theta_h$ at the current tax rates is sufficient to assess whether a small adjustment in salience will generate efficiency benefits. For example, if the government decides that it is going to raise total taxes on $x$ by a small amount, this formula provides guidance for selecting which one of the available tax instruments should be increased. In contrast, if the planned tax increase is large, policymakers should be cautious of relying on this formula because the change in the magnitudes of the taxes could induce changes in their salience.

3. Conclusion

A long literature within public finance considers how to minimize the efficiency cost of distortionary taxation. Motivated by new empirical findings that a tax’s salience affects consumer behavior, this paper explored how attention to salience can provide policymakers with an extra degree of freedom for reducing a commodity tax’s excess burden. More generally, the results illustrate that careful attention to decision-making biases may offer unexplored possibilities for improving consumer welfare through the manipulation of commonly-available (but frequently overlooked) policy tools.

Several limitations are important to keep in mind when interpreting the theoretical results presented here. Most importantly, the model here abstracts from considerations that may shape the optimal degree of tax salience in the real world. For example, when agents are heterogeneous in the extent to which they respond to a given tax – that is, when a single tax instrument has different salience for different decision-makers – it will not in general be possible to achieve the first-best, at least when all agents must face the same tax instruments. Nonetheless, it would be straightforward to generalize the approach described in Section 2 for making incremental adjustments in salience to settings characterized by such heterogeneity. Along the same lines, by focusing on the case of a representative consumer, I have ignored distributional effects from the choice between high- and low-salience taxes.
In reality, decision-makers may exhibit behavioral biases in ways that correlate with individual characteristics, and such patterns can have important implications for the design of policy. For example, when high- and low-income consumers differ in their attentiveness to low-salience taxes, governments can manipulate tax salience to reduce commodity tax regressivity (Goldin and Homonoff, 2013). Similarly, the salience of a tax may affect its incidence between consumers and producers (Chetty, Looney and Kroft, 2009).

Another limitation is that, contrary to what is assumed here, the government’s objective function might seek to avoid policies that would induce its citizens to make mistakes. If so, policymakers may not wish to implement the optimal combination of high- and low-salience taxes because doing so would induce taxpayers to (accidentally) depart from the allocation that would be privately optimal for them to consume. For further discussion of such issues, refer to Gamage and Shanske (2011) and Goldin (2012).

Finally, the results highlight several promising avenues for future research. The first is the desirability of new research into the factors that shape consumers’ attentiveness to a tax, and in particular, to the conditions that determine whether consumers will remain inattentive as the size of the tax increases. Such research would be beneficial given the efficiency-enhancing potential of a tax instrument that is “sustainably” low-salience – i.e., that remains low-salience even when levied at high rates – as discussed in Section 2. Second, the results suggest the need to reconsider accepted intuitions in the field regarding the proper role of commodity taxation. For example, the Atkinson-Stiglitz theorem stands for the proposition that commodity taxes are undesirable in the presence of a non-linear income tax, apart from special cases. However, when the government has multiple options for designing commodities taxes, and the options differ in their salience, the results here suggest that some role for commodity taxes may be optimal (at least when the income tax is fully-salient). Finally, it may be worthwhile to consider the implications of salience for the optimal allocation of taxes across commodities. The canonical Ramsey rule suggests levying taxes based on the elasticity of consumers’ demand for the taxed goods; the re-
sults here suggest the optimal policy depends on whether observed elasticities stem from inelastic preferences or from the imposed taxes having low salience.

Appendix

This Appendix derives the formula for incremental adjustments to the balance between high- and low-salience taxes that was discussed in Section 2. Recall that a revenue-neutral shift towards the low-salience tax is defined as a marginal increase in \( t_l \) along with whatever change in \( t_h \) is required to leave total revenue constant. Define \( \theta_i(t_h,t_l) = \frac{\partial x_i}{\partial t_i} \big|_{(t_h,t_l)} \) for \( i \in \{h,l\} \), i.e., the ratio of the tax and price derivatives evaluated at taxes \( t_h \) and \( t_l \).

**Proposition A.1** Starting at taxes \( t_h \) and \( t_l \), a revenue-neutral shift towards the low-salience tax is desirable if and only if \( \rho \theta_l(t_h,t_l) + (1 - \rho) \theta_h(t_h,t_l) > \theta^* \), where \( \rho \) and \( \theta^* \) are defined as in Proposition 2.

**Proof** Suppose that the high and low-salience taxes are set at \( t_h \) and \( t_l \). Tracking the derivation of Lemma 2, it is straightforward to show that a revenue neutral shift towards the low-salience tax is welfare improving if and only if

\[
U_x(x,y) - pU_y(x,y) > 0 \tag{1.26}
\]

The next steps apply a series of Taylor approximations to the quantities in (1.26).

First, note that

\[
U_i(x(t_h,t_l),y(t_h,t_l)) \approx U_i^0 + (x(t_h,t_l) - x(0,0)) \left( U_{xx}^0 - pU_{xy}^0 \right) - (t_h + t_l) x(t_h,t_l) U_{xy}^0 \tag{1.27}
\]

where \( U_{ij}^0 \equiv U_{ij}(x(0,0),y(0,0)) \).\(^{22}\) Similarly,

\(^{22}\)To reach this result, note that the consumer’s budget constraint implies \( y(t_h,t_l) - y(0,0) = -p \left( x(t_h,t_l) - x(0,0) \right) - (t_h + t_l) x(t_h,t_l) \). As is common in the literature, this approximation abstracts from third-order and higher terms.
\( U_y(x(t_h,t_l), y(t_h,t_l)) \approx U_y^0 + (x(t_h,t_l) - x(0,0)) \left( U_{yx}^0 - p U_{yy}^0 \right) - (t_h + t_l) x(t_h,t_l) U_{yy}^0 \) (1.28)

Next, we can approximate

\[ x(t_h,t_l) \approx x(0,0) + \frac{\partial x}{\partial p} (\theta_h t_h + \theta_l t_l) \] (1.29)

where \( \theta_h \) and \( \theta_l \) denote \( \theta_h(t_h,t_l) \) and \( \theta_l(t_h,t_l) \) respectively.

Substituting (1.27), (1.28), and (1.29) into (1.26) allows us to rewrite (1.26) as:

\[ \frac{\partial x}{\partial p} (\theta_h t_h + \theta_l t_l) \gamma_0 - (t_h + t_l) x(t_h,t_l) (U_{xy}^0 - p U_{yy}^0) > 0 \] (1.30)

where \( \gamma_0 \equiv U_{xx}^0 - 2p U_{yx}^0 + p^2 U_{yy}^0 \).

Next, totally differentiating the consumer’s budget constraint and first-order condition at the no-tax baseline with respect to \( I \) yields

\[ \frac{\partial x}{\partial p} \bigg|_{t_h=0} = -\frac{U_{yx} - p U_{yy}}{\gamma_0} \].

Substituting this identity into (1.30) and rearranging terms allows us to rewrite the condition as

\[ \frac{\partial x}{\partial p} (\theta_h t_h + \theta_l t_l) + (t_h + t_l) \frac{\partial x}{\partial I} < 0 \].

Finally, rewriting in terms of elasticities yields:

\[ \varepsilon_{x,p} (\theta_h t_h + \theta_l t_l) > (t_h + t_l) \eta_{x,I} \omega_x \]

where \( \varepsilon_{x,p}, \eta_{x,I}, \) and \( \omega_x \) are defined as in Proposition 2. Dividing both sides of the equation by \( \varepsilon_{x,p} (t_h + t_l) \) and applying the definitions of \( \rho \) and \( \theta^* \) yields the result. \( \blacksquare \)
Figure 1.1: Illustration of Result
Chapter 2

Smoke Gets in Your Eyes: Cigarette Tax Salience and Regressivity

(with Tatiana Homonoff)


**Abstract**

Recent evidence suggests consumers pay less attention to commodity taxes levied at the register than to taxes included in a good’s posted price. If this attention gap is larger for high-income consumers than for low-income consumers, policymakers can manipulate a tax’s regressivity by altering the fraction of the tax imposed at the register. We investigate income differences in attentiveness to cigarette taxes, exploiting state and time variation in cigarette excise and sales tax rates. Whereas all consumers respond to taxes that appear in cigarettes’ posted price, our results suggest that only low-income consumers respond to taxes levied at the register.

Should governments levy commodity taxes at the register or include them in a good’s posted price? Traditional approaches to the economics of taxation offer little guidance to policymakers choosing between the two tax types. Indeed, neoclassical theory suggests that this aspect of tax design – the choice between “posted” and “register” taxes – does not affect consumer welfare because consumers correctly compute and account for all taxes that will be assessed on a given transaction. However, a series of recent findings call that invariance prediction into doubt. For example, Chetty, Looney and Kroft (2009) (CLK) present compelling evidence that consumers pay more attention to goods’ posted prices than to register taxes because the former are more salient – consumers see the posted tax-inclusive price when making their purchasing decisions. Related empirical findings by Finkelstein (2009) and Cabral and Hoxby (2013) are also consistent with the hypothesis that the salience of a tax shapes the extent to which consumers perceive it. This line of research suggests that
the policy choice between posted and register taxes may not be as irrelevant as neoclassical theory predicts.

This paper investigates the distributional effects of the government’s choice between posted and register taxes. Part I considers the case in which consumers differ in their attentiveness to register taxes – that is, when only some consumers take register taxes into account when making purchasing decisions. Drawing on a stylized model of consumer behavior, we show how a revenue-neutral shift from posted to register taxes reduces the tax burden on attentive consumers, unambiguously improving the welfare of that group.

We then turn to a practical implication of this insight. A concern with many commodity taxes is that they are regressive – they constitute a proportionately greater burden for low-income taxpayers. However, if low-income consumers pay more attention to register taxes than high-income consumers do, policymakers can reduce a tax’s regressivity by adding it at the register instead of including it in the commodity’s posted price. Conversely, when low-income consumers are relatively less attentive to register taxes, reducing a tax’s salience will exacerbate its regressivity. Hence, knowing how consumers’ attentiveness to register taxes varies by income is essential for understanding the distribution of a tax’s burden.

Part II investigates that question empirically in the context of cigarette taxes. Cigarette purchases are typically subject to two types of taxes in the United States: an excise tax, which is included in the cigarette’s posted price, and a sales tax, which is added at the register. Drawing on individual survey data about cigarette consumption, we exploit state and time variation in cigarette sales and excise tax rates to estimate the relation between the two tax types and cigarette demand. We find that both high- and low-income consumers respond to changes in the cigarette excise tax, but that only low-income consumers respond to changes in the sales tax rate on cigarettes. Although the empirical results are not conclusive, they are consistent with the hypothesis that attentiveness to cigarette register taxes declines by income. In conjunction with the theoretical insights from Part I, our empirical
findings support the notion that a revenue-neutral shift from posted to register taxes could reduce the burden of the cigarette tax on low-income consumers.

Because the choice between register and posted taxes is a practical question that policymakers must confront, the lack of economic literature on the topic is surprising. Although the recent paper by Chetty, Looney, and Kroft (discussed above) provides important insights into the relative efficiency of posted and register taxes, our analysis builds on theirs by investigating how the choice between the two tax designs affects the distribution of the tax’s burden between consumers. In particular, the aggregate nature of their data preclude CLK from investigating heterogeneity in consumer attentiveness – our focus here. Moreover, the welfare analytic tools developed in CLK are geared toward assessing the efficiency of a tax in the context of a representative-agent, rather than a tax faced by heterogeneous consumers. To our knowledge, our paper is the first in the literature to investigate the link between the salience of a tax and the distribution of its burden across consumers.

Our paper also fits into a nascent behavioral literature investigating heterogeneity in the extent to which individuals depart from neoclassical models of decision-making. For example, Hall (2010) documents income differences in the mental accounting heuristics that individuals employ when making financial decisions and Bar-Gill and Warren (2008) present survey evidence suggesting that low-income consumers are more likely to make financial mistakes. Similarly, Mullainathan and Shafir (2009) argue that a number of behavioral phenomena affect the poor in distinctive ways because that group lacks many of the resources used by higher-income consumers to improve decision-making quality (such as access to financial advising). In a different context, Shue and Luttmer (2009) present evidence that low-income voters are particularly prone to accidentally selecting the wrong candidate when voting ballots are designed in confusing ways.23

23One theory that has been advanced to explain these findings is the notion of “cognitive depletion,” the idea that making complicated or high-stakes decisions can deplete individuals’ cognitive resources, worsening the quality of subsequent decisions they make. If low-income decision-makers must make more of these decisions throughout the day, they may exhibit a greater number of behavioral biases than do higher-income decision-makers. See Spears (2011) or Mullainathan and Shafir (2010).
Our paper contributes to this growing literature by exploring a particular context in which cognitive limitations faced by all decision-makers (e.g. bounded attention and computational abilities) affect high- and low-income consumers in distinctive ways. Most notably, whereas other studies have found deviations from optimal decision-making to be greatest for low-income decision-makers, we find the opposite. At least in the context of cigarette taxation, it appears that lower-income consumers do a better job of accounting for register taxes when making purchasing decisions. Apart from our empirical results, the theoretical framework we employ can be readily applied to other contexts in which agents differ in the extent to which they respond optimally to policy changes.

The paper is organized as follows. Part I constructs a stylized model of consumer behavior and uses it to analyze the welfare effects of a policy shift from posted to register taxes. The model takes as its starting point the assumption that consumers differ in their attentiveness to register taxes. Part II investigates that assumption empirically, in the context of cigarette taxation. In particular, we investigate whether high- and low-income consumers respond differently to cigarette register taxes, using those groups’ responsiveness to posted taxes on cigarettes as a baseline. Part III concludes.

I. Tax Salience and Distribution

Part I demonstrates that when consumers differ in their attentiveness to register taxes, the government’s choice between posted and register taxes affects the distribution of a tax’s burden. In particular, replacing a posted tax with a register tax increases total tax revenue because only attentive agents consider the full after-tax price when determining their demand for the taxed good. That extra revenue accommodates a reduction in the total tax rate, generating a positive welfare effect for attentive consumers. Inattentive consumers also benefit from the reduction in the total tax rate, but their welfare gains are offset by optimization error induced by the register tax.
A. Setup

Our modeling approach is similar to that employed in Chetty, Looney and Kroft (2007), except that we allow for heterogeneity in agents’ attentiveness to register taxes. Suppose that society is composed of two agents (A and B) who make consumption decisions between some good \( x \), and a composite of all other goods, \( y \). Good \( x \) is subject to both a register tax and a posted tax, whereas good \( y \) is left untaxed. Both agents pay attention to posted taxes when making their consumption decisions, but only A takes register taxes into account. B ignores the register tax when choosing how much \( x \) to consume, treating it as if it was zero. The agents share a utility function \( U(x, y) \), and both have budget constraints of the form

\[
BC_i : (p + t_p + t_r)x_i + y_i \leq M_i \tag{2.1}
\]

where the agent’s type is denoted by \( i \in \{A, B\} \), \( p \) is the pre-tax price of \( x \), \( t_p \) is the posted tax , \( t_r \) is the register tax , \( M \) is income, and the pre-tax price of \( y \) is normalized to one.

Consumption is determined in two steps. First, agents choose their intended consumption bundle according to their perceived budget constraint (\( \hat{BC}_i \)). A is attentive to the register tax, so her perceived budget constraint matches her true budget constraint, \( BC_A = \hat{BC}_A \). In contrast, B misperceives the register tax to be zero: \( \hat{BC}_B : (p + t_p)x_i + y_i \leq M_i \). The \((x, y)\) pair that maximizes utility subject to the agent’s perceived budget constraint is the intended consumption bundle \((\hat{x}_i, \hat{y}_i)\).\(^{24}\) Note that B’s intended consumption bundle will be infeasible when it fails to satisfy her true budget constraint.

Because the bundle that agents consume must ultimately be feasible, closing the model requires specifying the final consumption bundle for agents whose intended consumption bundle is infeasible. Because A chooses a feasible bundle to begin with, her final bundle always equals her intended bundle, \((x_A, y_A) \equiv (\hat{x}_A, \hat{y}_A)\). To pin down consumption for B, we assume that agents who over-spend on \( x \) reduce their expenditures on \( y \) by the amount

\(^{24}\)That is, \((\hat{x}_i, \hat{y}_i)\) satisfies \( \text{argmax } U(x_i, y_i) \) s.t. \( \hat{BC}_i \) holds.
that they overspent on \( x \). In our notation: \( x_B = \hat{x}_B \) and \( y_B = M_B - (p + t_p + t_r) \hat{x}_B \). This assumption is natural for the case in which \( y \) represents all goods other than \( x \) and agents make at least some of their consumption decisions after purchasing \( x \); consumers who accidentally overspend on \( x \) will have less income available to spend on their remaining purchases (which are all part of \( y \)).

We are now in a position to link consumer demand to the two tax types. Assume for now that production of \( x \) is governed by constant returns to scale technology and that the market for \( x \) is perfectly competitive, so that \( p \) is fixed at the (constant) marginal cost of \( x \). Holding the pre-tax price and agents’ income fixed, we can express demand as a function of the taxes, \( x_i = x_i(t_p, t_r) \) and \( y_i = y_i(t_p, t_r) \). For \( A \), final consumption always equals intended consumption, so demand corresponds to the solution of the standard utility maximization problem: \( (x_A, y_A) = \arg \max_{x,y} U(x,y) \) s.t. \( BC_A \). Because the tax rates do not enter the utility function directly and because they appear symmetrically in the budget constraint, \( A \)’s demand will depend only on the total tax rate – the portion of taxes included in the posted price does not matter. Hence we can write \( x_A(t_p, t_r) = x_A(t_p + t_r, 0) \), or \( x_A(t_p + t_r) \) for short. And similarly for \( y_A(t_p, t_r) = y_A(t_p + t_r, 0) \), or \( y_A(t_p + t_r) \) for short. Note that in accordance with the neoclassical model’s invariance prediction, we have \( \frac{\partial x_A}{\partial t_r} = \frac{\partial x_A}{\partial t_p} = \frac{\partial x_A}{\partial p} \) and \( \frac{\partial y_A}{\partial t_r} = \frac{\partial y_A}{\partial t_p} = \frac{\partial y_A}{\partial p} \).

Deriving \( B \)’s demand is complicated by the fact that her intended consumption departs from her final consumption whenever she faces a positive register tax. By assumption, all of the income \( B \) overspends on \( x \) comes out of intended expenditures on \( y \); hence \( B \)’s final consumption of \( x \) equals \( B \)’s intended consumption of \( x \): \( x_B(t_p, t_r) = \hat{x}_B(t_p, t_r) \) for

---

25Note that we are implicitly assuming that \( x \) is a small enough portion of total consumption that an agent’s intended consumption of \( x \) is never infeasible, even after taking the register tax into account.

26In principle, one could choose a different rule for mapping consumers’ sub-optimal decision-making into feasible consumption bundles. Chetty, Looney and Kroft (2007) identify three intuitive “budget adjustment rules”: the one that we employ, as well as two others. Appendix B demonstrates that the qualitative results in this section are robust to all three of those rules. More generally, the Appendix demonstrates that our main result holds as long as individuals who misperceive the price of \( x \) to be lower than it really is end up allocating more of their income to \( x \) and less of their income to \( y \) relative to the case in which they take the true after-tax price of \( x \) into account.
all values of $t_p$ and $t_r$. Moreover, because $B$’s intended consumption of $x$ is insensitive to register taxes, $\tilde{x}_B(t_p, t_r) = \tilde{x}_B(t_p, t'_r)$ for all $t_r$ and $t'_r$, it must also be the case that $B$’s final consumption of $x$ is insensitive to register taxes, $x_B(t_p, t_r) = x_B(t_p, t'_r)$ for all $t_r$ and $t'_r$. Consequently, we can write $B$’s final consumption of $x$ as a function of the posted tax alone: $x_B(t_p, t_r) = x_B(t_p)$. Finally, because $B$’s perceived budget constraint matches her true budget constraint in the special case that $t_r = 0$, we can conclude that $B$’s demand for $x$ under any non-zero register tax corresponds to $B$’s optimal demand for $x$ when the register tax is zero:

$$x_B(t_p, t_r) = x_B(t_p) = x_B^* (t_p, 0) \quad (2.2)$$

where $x_B^*$ represents $B$’s optimal consumption of $x$, i.e. the amount of $x$ that $B$ would choose if her perceived budget constraint were equal to her true budget constraint.27

By substituting (2.2) into $B$’s true budget constraint, we can solve for $B$’s final consumption of $y$:

$$y_B(t_p, t_r) = M_B - (p + t_p + t_r) x_B(t_p). \quad (2.3)$$

Note that in contrast to the neoclassical model, $B$ responds differently to the two types of taxes: $\frac{\partial y_B}{\partial t_p} = \frac{\partial y_B}{\partial p} < \frac{\partial y_B}{\partial t_r} = 0$ and $\frac{\partial y_B}{\partial t_r} = \frac{\partial y_B}{\partial p} > \frac{\partial y_B}{\partial t_r}$.

### B. The Role of Tax Policy

To incorporate tax policy into the model, consider a government that must raise a fixed amount of revenue, $\bar{R}$, from register and posted taxes on $x$. How does the government’s choice between register and posted taxes affect the well-being of the agents? In particular, the policy we consider is a revenue-neutral increase in the register tax – that is, an increase in the register tax coupled with a reduction in the posted tax by an amount that leaves total revenue unchanged (at $\bar{R}$). Let $R$ denote total revenue collected by taxes on $x$, so that $R(t_p, t_r) = (t_p + t_r) (x_A + x_B)$.

---

27That is, $x_B^*$ is the value of $x$ that solves $argmax_{(x,y)} U(x, y)$ s.t. $BC_B$ holds.
If both agents were fully attentive to both types of tax, a one-dollar increase in the register tax could accommodate a one-dollar reduction in the posted tax; changing the balance between register and posted taxes would not affect the combined tax rate necessary to raise a given amount of revenue. When some agents are inattentive, however, the demand reduction that typically accompanies a tax increase will be muted. As a result, an incremental increase in the posted tax will, all else equal, raise less revenue than an incremental increase in the register tax:

\[
\frac{\partial R}{\partial t_p} = (x_A + x_B) + (t_p + t_r) \left( \frac{\partial x_A}{\partial p} + \frac{\partial x_B}{\partial p} \right) < (x_A + x_B) + (t_p + t_r) \left( \frac{\partial x_A}{\partial p} + 0 \right) = \frac{\partial R}{\partial t_r}.
\]

The reduction in the posted tax associated with a revenue-neutral increase in the register tax can be found by totally differentiating the government’s budget constraint:

\[
\left. \frac{\partial t_p}{\partial t_r} \right|_R = -\frac{x_A + x_B + (t_p + t_r) \frac{\partial x_A}{\partial p}}{x_A + x_B + (t_p + t_r) \frac{\partial x_A}{\partial p} + (t_p + t_r) \frac{\partial x_B}{\partial p}}.
\] (2.4)

How does a revenue-neutral increase in the register tax affect the combined tax rate, \(t_p + t_r\)? The effect of the shift is given by \(\frac{d(t_p+t_r)}{dt_r} \bigg|_R = \frac{\partial t_p}{\partial t_r} \bigg|_R + 1\). Assuming that \(x\) is a normal good, (2.4) implies that \(\frac{\partial t_p}{\partial t_r} \bigg|_R < -1\).28 Consequently, a revenue-neutral increase in the register tax is associated with a net reduction in the combined tax rate, \(\frac{d(t_p+t_r)}{dt_r} \bigg|_R < 0\). Put differently, the government can maintain revenue-neutrality while reducing posted taxes more than one-for-one with each register tax increase.

C. Welfare Effects for Attentive Consumers

What are the welfare effects of such a revenue-neutral shift towards register taxes? First, consider the effect of the shift on A’s welfare. Indirect utility for A is given by \(V_A (t_p, t_r) = \)

\(^{28}\)Note that the denominator in (2.4) equals \(\frac{\partial R}{\partial p}\). Hence it is positive as long as demand for \(x\) is not so sensitive that raising the posted tax would actually decrease revenue, an assumption we maintain throughout.
\[ U(x_A(t_p, t_r), y_A(t_p, t_r)) \]. The welfare effect of the shift for \( A \) is thus:

\[
\frac{dV_A}{dt_r} \bigg|_R = \frac{\partial V_A}{\partial t_r} + \frac{\partial V_A}{\partial t_p} \frac{\partial t_p}{\partial t_r} = -U_y(x_A, y_A)x_A \left( \frac{d(t_p + t_r)}{dt_r} \right) .
\]

Equation (2.5) states that the welfare effect of a revenue-neutral shift towards register taxes for attentive consumers stems entirely from the effect of the shift on the after-tax price of \( x \). Increasing the register tax by one-dollar accommodates a reduction in the combined tax rate on \( x \) of \( \frac{d(t_p + t_r)}{dt_r} \). For each dollar that the after-tax price of \( x \) is reduced, \( A \) has \( x_A \) dollars more of income to spend on other goods \( y \). The greater the marginal utility of \( y \), the greater \( A \)'s welfare gain will be. Because we know that a revenue-neutral shift towards 

register taxes reduces the combined tax rate, \( \frac{d(t_p + t_r)}{dt_r} < 0 \), we can conclude that the shift unambiguously increases the welfare of the attentive agent.

Equation (2.5) is the main result of the section, and the intuition is straightforward. Replacing a posted tax with a register tax raises total revenue because only attentive agents reduce their demand for \( x \) in response to the higher after-tax price. The extra revenue accommodates a reduction in the combined tax rate on \( x \), generating a positive welfare effect for attentive consumers.

In addition to allowing us to sign the welfare effect of the policy for attentive consumers, (2.5) also highlights that the magnitude of the welfare effect depends upon the extent to which increases in the register tax accommodate revenue-neutral reductions in the posted tax, \( \frac{d(t_p + t_r)}{dt_r} \). With some algebra, we can rewrite (2.4) to obtain:

\[
\frac{d(t_p + t_r)}{dt_r} \bigg|_R = -\frac{\tau \varepsilon_B \phi_B}{1 - \tau (\varepsilon_A \phi_A + \varepsilon_B \phi_B)}
\]

Note that the envelope theorem applies here because \( A \)'s final consumption bundle matches her optimal consumption bundle.

This result can be weakened by endogenizing agents’ decisions about whether to pay attention to register taxes. If a small increase in the register tax causes a large number of formerly inattentive agents to start taking register taxes into account, the shift might necessitate an increase in the combined tax rate. In such cases, the shift to register taxes would actually generate a negative income effect, reducing the welfare of all agents. Consequently, the results presented here are most applicable to situations in which small changes in the tax rate do not induce dramatic shifts in which agents are attentive.
where \( \tau \equiv \frac{t_p + t_r}{p + t_p + t_r} \) is tax as a fraction of the tax-inclusive price of \( x \), \( \varepsilon_i \equiv -\frac{\partial x_i}{\partial p} \) is the own-price elasticity of demand for \( x_i \), and \( \phi_i \equiv \frac{x_i}{x_i + x_{-i}} \) is the fraction of \( x \) consumed by type \( i \).

The magnitude of the reduction in the combined tax rate permitted by a revenue-neutral shift thus depends upon three factors: the fraction of \( x \) consumed by inattentive consumers, the sensitivity of demand for \( x \), and taxes as a share of price. To understand the role of these factors, recall that inattentive consumers are the only ones who behave differently under the two taxes. The greater the fraction of \( x \) consumed by that group, the more important their inattentiveness will be in determining revenue from the taxes. Similarly, when \( B \)’s demand for \( x \) is highly elastic, the revenue advantage of a register tax over a posted tax is especially large – the posted tax causes a large amount of substitution away from the taxed good that the register tax avoids. Finally, the larger are taxes as a share of \( x \)’s price, the more that changes in those taxes affect consumer behavior for a given price-elasticity. Thus the welfare effect of a revenue-neutral shift towards register taxes is positive for \( A \), and increasing in \( \tau \), \( \varepsilon \), and \( \phi_B \).

**D. Welfare Effects for Inattentive Consumers**

What about the inattentive agent? The change in \( B \)’s indirect utility following a revenue-neutral shift towards register taxes is given by

\[
\frac{dV_B}{dt_r} \bigg|_R = U_x(x_B, y_B) \frac{\partial x_B}{\partial t_p} \bigg|_R + \frac{\partial y_B}{\partial t_r} \bigg|_R + U_y(x_B, y_B) \left( \frac{\partial y_B}{\partial t_p} \frac{\partial t_p}{\partial t_r} \bigg|_R + \frac{\partial y_B}{\partial t_r} \right)
\]

Differentiating \( B \)’s budget constraint with respect to \( t_p \) and \( t_r \) yields:

\[
\frac{\partial y_B}{\partial p} = -x - (p + t_p + t_r) \frac{\partial x_B}{\partial p} \quad \text{and} \quad \frac{\partial y_B}{\partial t_r} = -x.
\]
Substituting those conditions into the above expression gives the effect of the shift on \( B \)'s welfare:

\[
\frac{dV_B}{dt \mid_R} = -U_y(x_B, y_B) x_B \left( \frac{d(t_r + t_p)}{dt \mid_R} \right) + \frac{\partial t_p}{\partial t_r \mid_R} \frac{\partial x_B}{\partial p} \mu
\]

(2.7)

where \( \mu \equiv U_x(x_B, y_B) - (p + t_r + t_p) U_y(x_B, y_B) \).

From (2.7), we can see that the net welfare effect for inattentive consumers is ambiguous. The first term is strictly positive: like the attentive consumer, \( B \) benefits from the fact that the shift accommodates a reduction in the combined tax rate. In particular, a revenue-neutral shift that increases the register tax by $1 reduces the combined tax rate by \( \frac{d(t_p + t_r)}{dt \mid_R} x_B \) dollars of income. On the other hand, the second term in (2.7) is negative and reflects the fact that by raising the register tax, the shift pushes \( B \) further from her privately optimal consumption bundle. To understand the pieces of the term, note that a revenue-neutral shift that increases the register tax by one-dollar is associated with a posted tax reduction of \( \frac{\partial t_p}{\partial t_r \mid_R} \frac{\partial x_B}{\partial p} \), which prompts \( B \) to increase her consumption of \( x \) by \( \frac{\partial x_B}{\partial t_r \mid_R} \frac{\partial x_B}{\partial p} \) and reduce her consumption of \( y \) by \( -\frac{\partial t_p}{\partial t_r \mid_R} \frac{\partial x_B}{\partial p} (p + t_p + t_r) \).

If \( B \)'s consumption bundle were optimal, this substitution would not have any utility cost because the marginal utilities of expenditures on \( x \) and \( y \) would be equal. However, because \( B \) consumes too much \( x \) and too little \( y \) relative to the amounts that would be privately optimal given her true budget constraint, declining marginal utility in \( x \) and \( y \) implies that \( \mu = U_x(x_B, y_B) - (p + t_r + t_p) U_y(x_B, y_B) \leq 0 \). Thus a revenue-neutral increase in the register tax generates optimization error that reduces \( B \)'s utility by \( \frac{\partial t_p}{\partial t_r \mid_R} \frac{\partial x_B}{\partial p} \mu \). In general, either the positive or the negative welfare effect in (2.7) may dominate.

That even inattentive consumers can be made better off by a shift towards register taxes is somewhat surprising. The explanation lies in the fact that when the register tax is small, the utility cost of optimization error stemming from the register tax is small as well, but the positive welfare effect stemming from the lower combined tax rate can still be sizable. In particular, when the register tax is small, \( (x_B, y_B) \) will be close to \( (x^*_B, y^*_B) \) – the optimal

\[31\] That is, the standard first-order condition \( U_x(x_B, y_B) = (p + t_r + t_p) U_y(x_B, y_B) \) implies \( \mu = 0 \).
bundle in $B$’s true budget set. Consequently, the marginal utilities of expenditures on $x$ and $y$ will be close in size, implying a value of $\mu$ near zero.\footnote{Formalizing this intuition is straightforward. Assume that utility is additively separable in $x$ and $y$ so that $U(x,y) = u(x) + v(y)$. Then Taylor approximations of $\mu$ around $(x_B^*,y_B^*)$ and of $x_B$ around $x_B^*$ yield $\mu \approx -t_r \frac{\partial \mu}{\partial p} \left( U_{xx}(x_B^*) + (p + t_r + t_p)^2 U_{yy}(y_B^*) \right)$.} In contrast, the magnitude of the positive welfare effect in (2.7) depends on the level of the marginal utility of $y$, not the difference in the marginal utilities of $x$ and $y$; hence it stays positive even when $t_r \approx 0$. Thus when the register tax is small, revenue-neutral increases in $t_r$ tend to benefit both types of consumers. In the special case that $t_r = 0$, the optimization error associated with a small increase in $t_r$ is exactly zero, implying that a revenue-neutral shift towards register taxes always benefits inattentive consumers.\footnote{Because the shift also benefits attentive consumers, this result implies that the optimal register tax is always positive.}

To better understand the other factors that determine whether a shift will benefit inattentive consumers, we can substitute (2.4) into (2.7) and rearrange terms to obtain:

$$\frac{dV_B}{dt_r} \bigg|_R > 0 \iff \phi_B (U_y(x_B,y_B)(tp + t_r) + \mu) + \phi_A \mu (1 - \tau \varepsilon_A) > 0 \quad (2.8)$$

When $x$ is primarily consumed by attentive consumers, i.e. $\phi_A \approx 1$, (2.8) shows that revenue-neutral shifts towards register taxes tend to harm inattentive consumers.\footnote{Note that $\tau \varepsilon < 1$ follows from our maintained assumption that $\frac{\partial \mu}{\partial p} > 0$, i.e. that demand for the taxed good is not so sensitive that increasing the posted tax reduces revenue.} Intuitively, a revenue-neutral shift towards register taxes accommodates only a small reduction in the combined tax rate when most consumers are attentive because the revenue differences between the two tax types will be small. However, inattentive consumers still bear the full utility costs of optimization errors that stem from the higher register tax following the shift. In contrast, when $\phi_B \approx 1$, inattentive consumers benefit from a shift whenever $U_y(x_B,y_B)(tp + t_r) + \mu > 0$. Whether this condition holds depends on the relative welfare effects of the reduction in the combined tax rate and the optimization error induced by $t_r > 0$.\footnote{Because our focus in the rest of the paper is on heterogeneity between agents, we do not further develop the case in which all agents are inattentive. See Chapter 1 for further results.}
E. Summary and Extensions

In summary, while a revenue-neutral shift towards register taxes always benefits attentive consumers, the net welfare effect for inattentive consumers is ambiguous. Like A, B benefits from the lower combined tax on x associated with the shift. However, unlike A, B is driven by the shift to misallocate income between x and y (relative to the allocation that maximizes B’s private utility). When register taxes are small, the utility cost of that misallocation is small as well, and the positive welfare effect dominates. But when register taxes are large, the utility cost of the misallocation may be large as well. Additionally, when x is primarily consumed by attentive consumers, the positive welfare effects of the shift are muted for attentive and inattentive consumers alike.

For simplicity, we have assumed that the pre-tax price of x is fixed at $p$. In reality, firms may adjust the price they charge for x in response to changes in the type of tax imposed. If a shift from posted to register taxes induced firms to raise $p$ by a sufficient quantity, the policy could end up increasing the after-tax price of x, generating a negative welfare effect for all consumers.

Appendix A expands the model to the case of endogenous producer prices. We show that a revenue-neutral shift towards register taxes makes attentive consumers better off when supply of the taxed good is relatively elastic, in particular when $\varepsilon^S \tau > 1$, where $\varepsilon^S \equiv \frac{\partial s(p)}{p} \frac{(p+tp+tr)}{x_A+x_B}$ is the supply elasticity of x with respect to its after-tax price. In contrast, when $\varepsilon^S \tau < 1$, the reduction in $t_p$ caused by the shift is more than offset by an increase in $p$, resulting in a net increase in the after-tax price of x. Intuitively, when the supply of x is inelastic, the incidence of a posted tax falls on producers; reducing the posted tax and replacing it with a register tax – to which some consumers are less sensitive – allows producers to shift the incidence of the tax back on to consumers. Thus once one accounts for the endogeneity of producer prices, the welfare results presented in Part I apply only to goods for which demand is relatively inelastic and supply is relatively elastic – that is,
II. The Relation Between Cigarette Tax Attentiveness and Income

In Part I, we showed that policymakers can manipulate the salience of a tax to redistribute the tax’s burden between attentive and inattentive agents. In practice, policymakers are often concerned with how the burden of a tax is distributed by income. In particular, a concern with many commodity taxes is that they are regressive – that is, they constitute a disproportionately greater burden for low-income consumers. An implication of the results in Part I is that if the poor tend to pay more attention to register taxes than the rich, a shift towards register taxes will make a commodity tax more progressive. On the other hand, if low-income consumers are less attentive to register taxes, such a shift would exacerbate the tax’s regressivity. As such, it is important to determine whether attention to register taxes varies by income, and if so, whether high- or low-income consumers are the more attentive.

In Part II, we undertake that task in the context of cigarette taxes. There are good reasons to expect that low-income consumers will be more attentive to register taxes on cigarettes. In particular, the utility cost of optimization errors will tend to be greater for those with less income to spend on other goods. As a result, low-income consumers should be particularly motivated to spend the effort required to take register taxes into account. On the other hand, other factors could push high consumers to pay more attention to register taxes. For example, because the rich tend to consume more of each good, the magnitude of their optimization errors tends be greater as well. Appendix C utilizes a cognitive cost model to explore these tensions more formally. For the case of cigarettes, the analysis sug-

\[ \text{An important implication of this result concerns the case in which } e^{\xi} \tau < 1. \text{ For such goods, a revenue-neutral shift from posted to register taxes – the opposite of the policy considered above – will benefit both attentive and inattentive consumers. Attentive consumers benefit because the reduction in the pre-tax price more than offsets the increase in the total tax on } x, \text{ resulting in a net decrease in } x \text{’s after-tax price. Inattentive consumers also benefit from the after-tax price reduction, and because the shift is from register to posted taxes, it reduces the magnitude of their optimization error. Of course, the final incidence of either type of shift depends on the relative degree to which each type of consumer gains from producer surplus.} \]
gests that attentiveness to cigarette register taxes is likely to decline by income.\textsuperscript{37} However, because it is difficult to predict which group will be more attentive on the basis of theory alone, the remainder of Part II is primarily empirical.

Our goal is to investigate whether low-income cigarette consumers are more attentive to register taxes than high-income consumers are. Cigarette purchases are subjected to two types of tax in the United States: an excise tax, which consumers see reflected in the posted price, and a sales tax, which is typically added at the register. We use state and time variation in these tax rates to estimate how consumers respond to each type of tax. We assume that consumers fully account for posted taxes, so that inattention to register taxes can be measured by the gap between consumers’ responsiveness to register taxes and their responsiveness to posted taxes.

Part II is structured as follows. We begin by investigating whether the general population appears to pay more attention to register taxes than to posted taxes on cigarettes. The analysis applies the basic empirical strategy of CLK to a new product (cigarettes instead of beer) and at a different unit of observation (individual instead of aggregate consumption data). We then turn to our central question, whether attentiveness to cigarette register taxes differs by income, which we assess empirically by interacting the excise and sales tax variables with respondents’ income. Finally, we undertake a number of robustness tests to investigate whether our results actually reflect heterogeneous attentiveness to register taxes as opposed to various alternative explanations.

\textbf{A. Data}

We obtain cross-sectional micro data on cigarette consumption from the Behavioral Risk Factor Surveillance System (BRFSS), supported by the National Center for Chronic Dis-

\textsuperscript{37}The framework we develop does not make a uniform prediction for all goods, but rather highlights the factors that determine which income group will be more attentive to register taxes on a particular good. In general, high-income consumers tend to be less attentive to register taxes on goods, like cigarettes, for which demand is relatively insensitive to income.
ease Prevention and Health Promotion and the Centers for Disease Control and Prevention. The BRFSS is a state-based telephone survey system that tracks health conditions and risk behaviors of individuals 18 years and older. The number of states participating in the survey has grown over time, from 15 in 1984 to 50 in 1994 (as well as the District of Columbia).\textsuperscript{38}

We follow CLK by dropping two states from the analysis: Hawaii, because sales taxes in that state are included in the posted price, and West Virginia, because of frequent changes to that state’s sales tax base over the sample period. After dropping observations that are missing demographic variables, our final data set contains approximately 1.3 million observations. Because the survey disproportionately samples certain groups, we use weighted regressions to obtain representative estimates.

The BRFSS data contains two measures of smoking demand: whether the respondent is a smoker (smokes at least one cigarette every day) and how many cigarettes the respondent typically consumes each day. Although the BRFSS questionnaire asked respondents about smoking participation in each year of the survey, data on the number of cigarettes consumed are only available through 2000. Consequently, our analysis restricts the sample to those interviewed between 1984 and 2000.\textsuperscript{39} The BRFSS also collects information on a number of demographic variables, including income.\textsuperscript{40}

Data on state-level cigarette excise tax rates, sales tax rates, and average cigarette prices were obtained from the \textit{Tax Burden on Tobacco} 2008 report, published by Orzechowski and

\textsuperscript{38}To investigate whether the changing composition of states was biasing our results, we restricted our analysis to the 33 states that have been in the sample since 1987. The qualitative results were unchanged by that restriction.

\textsuperscript{39}Extending our empirical analysis through 2009 and using only the outcome variables available in the later years yields results similar to those obtained from our sample.

\textsuperscript{40}We make use of information concerning the respondent’s age, race, sex, educational attainment, marital status, employment status, and income. Household income is measured in terms of income-categories. Two problems arise when using this variable. First, the income measure is top-coded at a relatively low value ($75,000 for much of our sample period). Second, the income categories are not adjusted for inflation, making it difficult to compare respondents in the same category over time. Rather than attempt to convert the BRFSS income category data into a measure of real income, we measure income in percentile terms, assigning respondents the midpoint of the percentiles of their income category boundaries. For example, if 10 percent of the sample in one year reports an income between zero and $10,000, all individuals in that income category in that year are assigned a value of 0.05. This approach is similar to that employed by Franks et al. (2007).
Walker (and previously by the Tobacco Institute). We gathered information on the exact date of enactment of sales tax changes from a number of sources including the World Tax Database (University of Michigan), state government websites, and archives of local newspaper accounts. Following convention, our measure of state tax rates includes local taxes to the extent that they are uniform across the state.

Whereas the sales tax is an ad valorem tax (consumers are charged a fixed fraction of a good’s price), the excise tax is a unit tax (consumers pay a set dollar amount per pack, regardless of the pre-tax price). In order to make the two types of taxes comparable for the empirical analysis, we convert the excise tax to an ad valorem tax using the method described in CLK.41

Both sales and excise taxes increased between 1984 and 2000 (Figures 2.1a and 2.1b). In 1984, 38 states imposed sales taxes on cigarettes, and the median sales tax rate was 4 percent. By 2000, 45 states imposed sales taxes on cigarettes, and the median sales tax rate had climbed to 5 percent. Similarly, median state excise taxes on cigarettes increased from 14 cents in 1984 to 34 cents in 2000. In addition, the federal excise tax on cigarettes increased three times over the same period, climbing from 16 to 34 cents per pack. Table 2.1 presents summary statistics on U.S. cigarette taxation.

Figure 2.2 shows that aggregate cigarette consumption in the United States declined between 1984 and 2000. That decline, however, was not uniform across the population. Figure 2.3 separately plots smoking participation rates over time for the highest and lowest income quartiles. Low-income individuals were more likely to smoke than high-income ones in 1984, and that gap widened over time. Smoking demand measures are summarized in Table 2.2.

41We divide the excise tax by the average national price of a pack of cigarettes in 2000, adjusted for inflation. The rationale for using the inflation-adjusted national price in 2000 as a proxy for the true price is to avoid endogeneity problems arising from the fact that changes in cigarette prices are likely correlated with unobserved shocks to smoking demand.
B. Attentiveness to Cigarette Taxes in the General Population

We begin our empirical analysis by investigating whether consumers in the general population respond differently to register and posted taxes on cigarettes. The neoclassical model predicts that the salience of a tax (e.g., whether it is included in the posted price or added at the register) should not affect how consumers respond to it. To see this formally, suppose that demand for a good \( x \) depends on a consumer’s income \( I \) and the price of \( x \), \( p_x \):

\[
x = x(p_x, I).
\]

Purchases of \( x \) are subject to both a sales tax and an excise tax, so that the final price of \( x \) is given by \( p_x = p(1 + t)(1 + s) \), in which \( p \) is the pre-tax price of \( x \), \( t \) is the excise tax rate, and \( s \) is the sales tax rate.\(^{42}\)

Because the excise and sales tax affect the price of \( x \) symmetrically, we have that

\[
\frac{\partial x}{\partial \log(1 + t)} = \frac{\partial x}{\partial \log p_x} \frac{\partial \log p_x}{\partial \log(1 + t)} = \frac{\partial x}{\partial \log p_x} \frac{\partial \log p_x}{\partial \log(1 + s)} = \frac{\partial x}{\partial \log(1 + s)}
\]

In words, how consumers adjust their demand for \( x \) in response to a tax change should not depend on whether the change occurred in the excise tax rate or the sales tax rate.\(^{43}\)

CLK assess this prediction for the case of beer by linking changes in aggregate beer consumption by state to changes in the state’s sales tax rate and excise tax on beer. They find that changes in the beer excise tax are negatively and significantly correlated with changes in beer consumption, whereas sales tax changes appear to have little effect. As a result, CLK conclude that the neoclassical model is mistaken and that the salience of a tax affects how consumers respond. Because they lack disaggregated consumption data, CLK are unable to assess whether the salience of a tax affects different parts of the population differently, our goal in Section II.C.

\(^{42}\)Some states do not include the excise tax in the price used to calculate the sales tax, so that final prices are given by \( p_x = p(1 + t + s) \). Because the excise and sales tax still affect the price of \( x \) symmetrically in such states, the neoclassical model predicts that demand should respond identically to sales and excise tax changes of the same proportion.

\(^{43}\)Two assumptions are important for this result: first, that tax rates only enter consumer utility through their effect on product prices, and second, that \( p_x \) is the only price that affects demand for \( x \). We maintain the first assumption throughout but consider the implications of relaxing the second in Section II.E.
Our analysis in this section differs from CLK by focusing on cigarettes instead of beer and by using individual survey data rather than aggregate state consumption data. Our baseline empirical model takes the form:

\[
y_{ismt} = \alpha + \beta_1 \tau_{smt} + \beta_2 \tau_{smt} + \gamma x_{smt} + \delta z_{ismt} + \mu_s + \lambda_t + \pi_m + \epsilon_{ismt} \tag{2.9}
\]

where the unit of observation is an individual \(i\) in state \(s\), calendar month \(m\), and year \(t\). The dependent variable \(y\) represents cigarette demand, \(\tau^e\) is the log excise tax rate, \(\tau^s\) is the log sales tax rate, \(x\) are covariates that do not vary between individuals interviewed in the same state, month, and year, and \(z\) are individual-level covariates. We include state fixed effects \(\mu_s\) to capture unobserved factors that are correlated with both state tax rates and the level of smoking demand. Year fixed effects \(\lambda_t\) capture time trends in smoking demand as well as yearly shocks to national cigarette consumption, such as a national anti-smoking campaign. Finally, \(\pi_m\) is a calendar month effect, which accounts for seasonal or monthly patterns in cigarette demand.

As is standard in the cigarette demand literature,\(^{44}\) we model the decision of whether an individual smokes (the extensive margin) separately from the decision of how much to smoke, conditional on being a smoker (the intensive margin). Consequently, in some specifications \(y\) is a binary choice variable indicating whether the individual reports being a smoker, and in other specifications \(y\) is the non-zero count of the number of cigarettes consumed in the last month, where the sample is restricted to self-reported smokers. This “double-hurdle” model is common in the cigarette demand literature because the decision of whether to smoke may be fundamentally different than the decision of how much to smoke, and is informative as to whether taxes affect consumption by turning smokers into non-smokers or by inducing current smokers to reduce the number of cigarettes they smoke.\(^{45}\)

\(^{44}\)See Chaloupka and Warner (2000) for a helpful review of the extensive literature on estimating cigarette demand.

\(^{45}\)A drawback of the two-part approach is that estimation results for the intensive margin may be biased by
Table 2.3 presents the results of this analysis. The specifications in Columns 1 and 4 regress smoking demand on the two tax rates, individual demographic variables, and state, year, and calendar month fixed-effects. Since state taxes are often increased to meet budgetary shortfalls in bad economic times, it is likely that tax rate changes are correlated with state-level economic variables that are not captured by state fixed effects. If cigarette consumption is also correlated with the business cycle, this omitted variable could bias our results. To account for this possibility, Columns 2 and 5 include state-level measures of real income and unemployment rate.

Columns 3 and 6 add an interaction between income and a linear time trend. To motivate this addition, recall that smoking participation rates fell more steeply over the sample period for higher income consumers (Figure 2.3). Although this decline might stem from rising tax rates over the sample period, it could also reflect a secular trend in smoking consumption at the top of the income spectrum, such as a shift in cultural attitudes about smoking among high SES individuals. Because tax rates trend upwards over the sample period, a secular trend in smoking demand among high-income consumers could be conflated with the two tax-income interaction terms in the regression. The inclusion of the time trend in Columns 3 and 6 accounts for this possibility.

The regressions in Table 2.3 show the effect of taxes on the intensive and extensive margins separately. In order to provide a better picture of the overall effect of a tax change on cigarette demand, Table 2.4 follows the procedure laid out in McDonald and Moffitt (1980) to combine the intensive and extensive margin estimates. In particular, one can decompose the conditional expectation of cigarette demand into its intensive and extensive components:

\[
E[y|x] = E[y|x, y > 0] \times P(y > 0|x),
\]

changes to the composition of the smoking population. We investigate the robustness of this specification in Section II.E.2.

We estimate demand on the extensive margin with a linear probability model. A Probit model yields similar results. Because unobserved shocks to smoking demand may be correlated across time for consumers living in the same state, all tables report standard errors that are clustered at the state level.

Real state income data comes from the Bureau of Economic Analysis and the state unemployment rate data comes from the Bureau of Labor Statistics. Both variables are measured quarterly.
where $y$ represents cigarette demand and $x$ represents the covariates. Using the product rule, the total effect of a change in one of the covariates on cigarette demand is given by:

$$
\frac{\partial E[y|x]}{\partial x} = \frac{\partial E[y|x, y > 0]}{\partial x} \cdot P(y > 0|x) + \frac{\partial P(y > 0|x)}{\partial x} \cdot E[y|x, y > 0].
$$

By utilizing sample estimates of $P(y > 0|x)$ and $E[y|x, y > 0]$, evaluated at the sample mean of each covariate, we can combine the estimated coefficients from the intensive and extensive margin regressions into a rough estimate of the overall effect of the taxes on cigarette demand.\(^{48}\)

The results in Tables 2.3 and 2.4 are consistent with a salience effect on the intensive margin: under our preferred specification, a one-percent increase in the cigarette excise tax is associated with a 0.34 percent reduction in cigarettes per month among smokers, whereas the point estimate on the sales tax term is close to zero and is not statistically significant. However, the coefficient on the sales tax is measured imprecisely, and consequently, we cannot reject the null hypothesis of equality between the two coefficients. On the extensive margin, the point estimate of the sales tax is slightly greater in magnitude than that of the excise tax, although here too the difference is not statistically significant.\(^{49}\) The coefficients on the excise tax estimates imply price elasticities of -0.52 on the extensive margin, -0.31 on the intensive margin, and -0.87 for combined demand. For the sales tax, the implied price elasticities are -0.32 on the extensive margin, -0.02 on the intensive margin, and -0.32 for overall cigarette demand.\(^{50}\) Overall, the evidence is inconclusive regarding the presence

\(^{48}\)When calculating standard errors for the aggregate effect, we ignore uncertainty in the sample averages of $P(y > 0|x)$ and $E[y|x, y > 0]$. This approximation is reasonable because the size of our sample guarantees those quantities are estimated precisely.

\(^{49}\)One complication also confronted by CLK is that the simple comparison between estimated tax coefficients can be misleading as a test of salience if the two tax types are passed through to consumers at different rates – that is, if $\frac{\partial (p + tp + tr)}{\partial t} \neq \frac{\partial (p + tp + tr)}{\partial t}$. Although a finding of differential pass-through is consistent with tax salience – see CLK pp. 1167-69 – it could also arise solely from differences in the two tax bases. As explained in Section E.1 below, we address this issue by comparing the sales tax coefficient with the effect of the pre-sales tax price of $x$, instrumented with the excise tax. For the general population analysis, an IV approach differs from Table 2.3 in that the estimated excise tax coefficient becomes greater in magnitude than the estimated effect of the sales tax, consistent with a salience effect. However, for both margins, the difference between the estimated coefficients remains statistically insignificant.

\(^{50}\)To compute the price elasticity implied by a tax coefficient, one must scale the coefficient by the rate
of a salience effect for the general population.

C. Attentiveness to Cigarette Taxes by Income

The inconclusive results for the general population in Section II.B might mask heterogeneous responsiveness across income groups. We now turn to our primary question of interest, whether low-income consumers are particularly attentive to cigarette register taxes. The baseline empirical model for this section is given by:

\[
y_{ismt} = \alpha + \beta_1 \tau_{esmt}^e + \beta_2 \tau_{ssmt}^s + \rho_1 \tau_{esmt}^e LI_{ismt} + \rho_2 \tau_{ssmt}^s LI_{ismt} + \eta LI_{ismt} + \\
\gamma x_{smt} + \delta z_{ismt} + \mu_t + \lambda_t + \pi_m + \epsilon_{ismt} (2.10)
\]

where \( LI \) is a binary variable indicating whether the respondent is low-income, defined as having income below the 25th percentile. Compared to the econometric model in II.2, this specification adds interaction terms between low-income status and the two tax rate variables.\(^5\) The coefficients on the two tax types, \( \beta_1 \) and \( \beta_2 \), describe how high-income consumers modify their demand in response to changes in the excise and sales taxes, respectively. In turn, the coefficients on the income-interaction terms, \( \rho_1 \) and \( \rho_2 \), measure whether low-income consumers are more or less sensitive to changes in the two tax types.

Our primary question is whether attentiveness to the sales tax varies by income. In answering this question, one must distinguish between attentiveness – the extent to which consumers account for a tax when making their consumption decisions – and price-sensitivity – which describes how a tax that consumers account for affects their optimal purchase. The pass-through rate may be obtained from Table 2.11. The estimated excise tax elasticities we find are on the larger side of those typically reported in the smoking literature. For example, Chaloupka and Warner (2000) report that recent estimates of (overall) cigarette demand range from elasticities of -0.14 to -1.23, but that most fall in the narrower range of -0.3 to -0.5. Gruber and Koszegi (2004) find an implied excise tax elasticity of -0.66 using the Consumer Expenditure Survey. Sales tax elasticities are not typically estimated in the smoking literature.

\(^{5}\)In addition to the main effect for low-income status, the individual demographics vector \( z \) also includes a continuous measure of income.
sales*low-income interaction term ($\rho_2$) may reflect differences in attentiveness between high- and low-income consumers, but it may also reflect differences in price-sensitivity by income. That is, a negative coefficient on $\rho_2$ could stem from high-income smokers being less sensitive to cigarette prices in any form, even if high- and low-income smokers were equally attentive to the sales tax.

To deal with this possibility, it is useful to introduce the notion of the “attention gap,” the amount by which a consumer’s responsiveness to the excise tax exceeds her responsiveness to the sales tax. For high-income consumers, the estimated attention gap is simply $\beta_2 - \beta_1$. In turn, for low-income consumers, the estimated attention gap is given by $(\beta_2 + \rho_2) - (\beta_1 + \rho_1)$. Recall that the neoclassical model described above predicts that consumers should respond identically to excise and sales taxes that they take into account. Consequently, we interpret a non-zero value of the attention gap as evidence that consumers account for one type of tax more than the other.

Although the sign and magnitude of the attention gap for a particular income group are interesting in their own right, more relevant to our analysis are changes in the attention gap by income. That is, we are less concerned with whether a particular group of consumers pays more attention to the excise tax relative to the sales tax, and more concerned with whether low-income consumers pay more attention to the sales tax (relative to the excise tax) than high-income consumers do.$^{52}$ It is easy to see that the estimated difference in attentiveness between high- and low-income consumers is given by:

$$\Delta\text{AttentionGap} = [(\beta_2 + \rho_2) - (\beta_1 + \rho_1)] - [\beta_2 - \beta_1] = \rho_2 - \rho_1$$

(2.11)

Intuitively, the sales*low-income coefficient ($\rho_2$) reflects changing responsiveness to

---

$^{52}$After all, it is the differences in behavior between high- and low-income consumers that shapes the distribution of a tax’s burden.
the sales tax by income, and the excise*low-income coefficient ($\rho_1$) removes the portion of that change due to changes in consumers’ price sensitivity. Hence, the gap between the coefficients on the two interaction terms rates, $\rho_2 - \rho_1$, measures the extent to which attentiveness to the register tax changes as income rises. When $\rho_2 - \rho_1 < 0$, high-income consumers pay less attention to the sales tax (relative to the excise tax) than low-income consumers do.

Table 2.5 presents our results. Columns 1 and 4 include the two tax rates, on their own and interacted with income. In addition, the regressions include demographic variables as well as state, year, and month fixed-effects. As before, Columns 2 and 5 add real state income and the state unemployment rate, and Columns 3 and 6 include an interaction between income and a linear time trend to capture the changing relationship between income and smoking behavior over time. The estimated coefficients on the demographic and macroeconomic variables are qualitatively similar to those reported in Table 2.3, and are omitted. Table 2.6 combines the intensive and extensive margin estimates into an overall effect, using the method described in Section II.B.\textsuperscript{53}

The results in Tables 2.5 and 2.6 are consistent with the theory that attentiveness to register taxes declines with income. As before, Columns 3 and 6 are our preferred specification.\textsuperscript{54} On both the intensive and extensive margins, the estimated tax coefficients suggest that high-income consumers respond less negatively to the sales tax than to the excise tax. The excise tax coefficients are negative and statistically significant, whereas the sales tax coefficients are statistically indistinguishable from zero.\textsuperscript{55} An F-test suggests that the difference in magnitude between the high-income tax coefficients is statistically significant on both margins.

For low-income consumers, the results paint a dramatically different picture. The coef-

\textsuperscript{53}The robustness checks that follow use the specification in Columns 3 and 6 as their baseline.
\textsuperscript{54}The only qualitative difference between specifications is the coefficient on the excise*low-income interaction, which declines sharply in magnitude once the income time trend is added to the model.
\textsuperscript{55}The high-income consumer price elasticities implied by these estimates are -0.61 (excise) and -0.06 (sales) on the extensive margin, and -0.31 (excise) and 0.18 (sales) on the intensive margin.
cicient on the interaction between low-income status and the sales tax and is negative and significant, implying that an increase in the sale tax is associated with a larger reduction in demand for low-income consumers than for high-income consumers. The small coefficient on the excise×low-income interaction term suggests that the result reflects a difference in attentiveness rather than a mere difference in price-sensitivity by income. 56

Recall from (2.11) that changes in the attention gap by income are captured by $\rho_2 - \rho_1$. Hence, to investigate whether low-income consumers are particularly attentive to register taxes, we test whether $\rho_1 = \rho_2$. The associated F-tests are reported in Tables 2.5 and 2.6. Under our preferred specifications, the F-statistics for the extensive and intensive margins are 8.60 and 5.14, respectively. Hence, our results are consistent with the hypothesis that low-income consumers pay more attention to cigarette register taxes than do high-income consumers. 57

So far, we have divided the analysis into low-income consumers on the one hand (those below the 25th percentile in income) and medium- to high-income consumers on the other. Although that aggregation is convenient for exposition, it may mask differences in attentiveness between medium- and high-income consumers. The regressions in Tables 2.7 and 2.8 introduce additional flexibility into the model by allowing consumers in each income

56 The low-income consumer price elasticities implied by these estimates are -0.30 (excise) and -1.13 (sales) on the extensive margin, and -0.30 (excise) and -0.59 (sales) on the intensive margin. One interesting result is that on both margins, the point estimate of the sales tax is more negative than point estimate of the excise tax for low-income consumers, although the difference is only significant on the intensive margin. This result could stem from differences in the goods included in the excise and sales tax bases, a possibility explored in Section II.D. Of course, it is also possible that the estimated sales tax coefficient is biased downward due to some omitted variable. However, unless that omitted variable was differentially correlated with smoking demand by high- and low-income consumers, it would not drive the differences in sales tax responsiveness that we observe.

57 Although the results from both margins are consistent with low-income consumers being more attentive than high-income consumers, several features of the analysis make the intensive margin results less convincing than those from the extensive margin. In particular, our finding that responsiveness to the two tax types varies by income on the extensive margin suggests the possibility that selection effects may confound our comparison of responsiveness on the intensive margin. Additionally, the positive point-estimate of the sales tax coefficient for high-income consumers, although not close to statistically significant, may indicate the presence of a selection effect or some other form of bias. A positive sales tax effect could also arise if the other goods in the sales tax base were strong substitutes for cigarettes; this possibility would bias our results if the substitution patterns between cigarettes and the other covered goods differed for high- and low-income consumers, a possibility explored in Appendix D.
quartile to respond to the taxes in different ways. The resulting specification is given by

$$y_{ismt} = \alpha + \beta_1 \tau_{esmt} + \beta_2 \tau_{smt} + \sum_{j \in \{II,III,IV\}} \left\{ \eta^j Q^j_{ismt} + \rho^j_1 \tau_{esmt} Q^j_{ismt} + \rho^j_2 \tau_{smt} Q^j_{ismt} \right\} + \gamma x_{smt} + \delta z_{ismt} + \mu_s + \lambda_t + \pi_m + \varepsilon_{ismt}$$ (2.12)

where $Q^j_{ismt}$ indicates whether consumer $i$ falls into income quartile $j$.

As before, we find that income differences in how consumers respond to the excise tax tend to be small and statistically insignificant. In contrast, responsiveness to the sales tax declines monotonically with income. F-tests for the equality of the attention gap between consumers in different income quartiles are reported in Tables 2.7 and 2.8 as well. The results suggest that attentiveness to cigarette register taxes declines monotonically by income.

**D. Tax Base Differences Between the Excise and Sales Tax**

Our strategy for measuring attentiveness has been to compare consumer responsiveness to excise and sales tax rates. When demand for cigarettes depends only on the price of cigarettes and income, any gap between how consumers respond to the sales tax and how they respond to the excise tax implies a departure from the neoclassical model (as explained

---

58 The results are similar when we include income as a linear interaction with the tax rates, or use the 20th or 30th income percentile to define the low-income group.

59 An implicit assumption in our analysis (and throughout the smoking literature) is that changes in cigarette taxes are uncorrelated with unobserved shocks to individuals' cigarette consumption. However, cigarette taxes are not set randomly; a positive shock to cigarette demand might prompt state legislators to raise excise taxes to capture additional revenue. Although such correlations could provide an alternative explanation for the discrepancy between the excise and sales tax coefficients in Section II.B, it is more difficult to imagine them driving the heterogeneous attentiveness results in Section II.C. That is, although there are many possible reasons for cigarette taxes to be correlated with unobserved shocks to smoking demand, there are fewer plausible reasons why adoption of such laws would be differently correlated with shocks to cigarette demand for high and low-income consumers. Moreover, to the extent that policymakers do consider cigarette demand by high- and low-income consumers differently when setting tax rates, it would be surprising if they took such behavior into account when setting the sales tax (for which cigarette sales constitute only a small fraction of total revenue). So although it appears unlikely that the endogenous adoption of tax laws is driving our main results, we cannot rule that possibility out definitively.
in Section II.B). In reality, the price of goods other than cigarettes may enter the cigarette demand function as well; if some of those other goods are also covered by the sales tax, the effect of a sales tax increase on cigarette demand will differ from the effect of an excise tax increase. This observation complicates our analysis because income differences in the attention gap may be due to differences in the excise and sales tax bases, rather than differences in attentiveness.

To clarify the nature of the problem, it will be helpful to discuss this tax-base effect in some detail. Under the neoclassical model, a tax can affect cigarette demand in two ways: by raising the price of cigarettes (a direct effect), and by raising the price of other goods (an indirect effect). Because the excise tax applies only to cigarettes, it generates only a direct effect. In contrast, the sales tax applies to many goods, and consequently, it generates both a direct effect and an indirect effect on cigarette consumption. As a result, income differences in the attention gap could reflect both income differences in attentiveness as well as income differences in the nature of the sales tax’s indirect effect. In particular, if the indirect effect of the sales tax on cigarette demand were more negative for low-income consumers than for high-income ones, it could be that a tax base effect rather than changing attentiveness is driving our results. That is, low-income consumers’ greater responsiveness to the sales tax could stem from income differences in how consumers adjust cigarette demand in response to price changes on other sales-taxed good.

Might income differences in the indirect effect of the sales tax be driving our results? It is difficult to dismiss this possibility out of hand. The indirect effect of the sales tax can be decomposed into an income effect and a substitution effect. By raising the price of many goods at once, the sales tax diminishes consumers’ purchasing power, causing them to reduce their consumption of cigarettes (the income effect). In addition, raising the price of other goods might cause consumers to substitute toward or away from cigarettes, depending on whether the other goods covered by the sales tax are primarily substitutes or

---

60 Approximately 40 percent of retail sales, according to CLK.
complements to cigarettes (the substitution effect). In theory, either of these effects could be more negative for low-income consumers. For example, the other sales-taxed goods could be important substitutes with cigarettes for well-off consumers, but not for low-income consumers. Similarly, the loss in real income associated with a sales tax increase could induce a bigger reduction in cigarette demand for low-income consumers.

Although we are unable to reject the possibility, we present two pieces of evidence that tax base effects are not responsible for all of the observed differences in consumer behavior by income. Our first check is motivated by the fact that some states impose a general sales tax, but exempt cigarettes from it.\textsuperscript{61} In states that exempt cigarettes from the sales tax, changes in the sales tax rate would not directly affect the price of cigarettes; the sales tax would not have a direct effect on cigarette consumption. However, sales tax changes would still affect the price of other sales tax-eligible goods. Hence, the indirect effect of the sales tax would still occur. Consequently, analyzing the effect of the sales tax in cigarette-exempting states allows us to measure income differences in the indirect effect of the sales tax.

If indirect effects were responsible for the observed differences in responsiveness to the sales tax by income, responsive to the sales tax should decline by income as much in states that exempt as in states that do not. Table 2.9 compares the effect of the sales tax in states that exempt cigarettes from the sales tax (“exempt states”) with the effect of the sales tax in states that include cigarettes in the sales tax base (“non-exempt states”). To do so, we modify our econometric model to allow heterogeneity in the effect of the sales tax between exempt and non-exempt states:

\[
y_{ismt} = \alpha + \beta_1 \tau_{smt} + \beta_2 \tau_{smt}E_{s} + \beta_3 \tau_{smt}(1 - E_{s}) + \\
\rho_1 \tau_{smt}LI_{ismt} + \rho_2 \tau_{smt}LI_{ismt}E_{s} + \rho_3 \tau_{smt}LI_{ismt}(1 - E_{s}) + \\
\phi E_{st} + \eta LI_{ismt} + \gamma x_{smt} + \delta z_{ismt} + \mu_{s} + \lambda_{t} + \pi_{m} + \epsilon_{ismt}
\]  

\textsuperscript{(2.13)}

\textsuperscript{61}In our sample, seven states exempt cigarettes from the sales tax base for at least one year.
where $E_{st}$ indicates whether state $s$ exempts cigarettes from the sales tax base in year $t$.

Table 2.9 presents the results of this analysis. In all specifications, the small number of state-year cells in the exempt category makes inference difficult. Columns 1 - 3 show that the effect of the sales tax on cigarette demand appears to vary substantially more by income in non-exempt states than in exempt states. The estimated sales*low-income coefficient in exempt states, $\rho_2$, is small in size and is statistically insignificant on both the extensive and intensive margins. In contrast, the sales*low-income coefficient in the non-exempt states, $\rho_3$, remains large and statistically significant. On the extensive margin, we are able to reject the hypothesis that the sales*low-income coefficient in the exempt states is as large as in the non-exempt states. A concern with these specifications is that high- and low-income consumers may exhibit different smoking behavior in exempt versus non-exempt state-years, independent of the sales tax. To address this possibility, Columns 4 - 6 introduce an interaction for low-income*exempt. On the extensive margin, Column 4 shows that the estimated effect of the sales tax on low-income consumers is only slightly more negative in non-exempt versus exempt states. In contrast, the intensive margin results in Column 5 are similar to those reported in Column 2: the point estimate of the sales*low-income coefficient is substantially more negative in exempt states, but the large standard errors on the sales*low-income coefficient for exempt states make conclusions of statistical significance impossible.

We also present a second check that tax base effects are not driving our results. Tax base effects are most likely to dampen the impact of the sales tax relative to the excise tax when the excise tax exempts important substitutes for cigarettes. Because other tobacco products constitute likely substitutes for cigarettes, there is less potential for tax base differences to play a role in states where the excise tax also applies to other tobacco products. Consequently, we restrict the analysis in Section II.C to states that apply the excise tax to

---

62 The low-income*exempt interaction is not statistically significant in any of the specifications.
63 For example, raising the excise tax might reduce cigarette demand substantially by inducing cigarette smokers to switch to cigars. In contrast, raising the sales tax would raise the price of both cigarettes and cigars, dampening the effect on cigarette consumption.
cigars and smokeless tobacco. Table 2.10 shows that the difference in sales tax responsiveness between high- and low-income consumers persists after restricting the sample to those states.

In summary, there are plausible reasons to believe that differences in the excise and sales tax bases could generate results similar to those presented in Section II.C. However, the evidence in Tables 2.9 and 2.10, while not conclusive, suggests that tax base effects cannot fully supplant attentiveness as an explanation for the large differences in behavior that we find between high- and low-income consumers.

E. Robustness Checks

1. Including Pre-Tax Prices in the Regression

One variable not included in our basic econometric model is the pre-tax price of cigarettes. On the one hand, the pre-tax price depends on both supply and demand; including it as a regressor could bias our results if it were correlated with unobserved shocks to consumer demand (the classic simultaneous systems problem). On the other hand, the pre-tax price enters the consumer’s demand function symmetrically with the excise and sales tax rates; excluding it from the regression may create an omitted variable bias if pre-tax price fluctuations were not equally correlated with the two tax types for high- and low-income consumers.\textsuperscript{64}

In this section, we modify our empirical strategy to account for the pre-tax price of cigarettes. Whereas previously we compared sales tax changes to excise tax changes, we now compare sales tax changes to changes in the posted price of cigarettes (the pre-tax price plus the excise tax). As before, this approach isolates income differences in attentiveness.

\textsuperscript{64}For example, Harding, Leibtag and Lovenheim (Forthcoming) find that excise taxes are passed through differently to high- and low-income consumers.
rather than changing price-sensitivity. The econometric model takes the following form:

\[ y_{ismt} = \alpha + \beta_1 p p_{smt} + \beta_2 \tau_{smt}^e + \rho_1 p p_{smt} LI_{ismt} + \rho_2 \tau_{smt}^e LI_{ismt} + \gamma x_{smt} + \delta z_{ismt} + \mu_s + \lambda_t + \pi_m + \epsilon_{ismt} \]  

(2.14)

where \( pp \) represents the (excise tax inclusive) log posted price of cigarettes. To address the possible correlation between pre-tax prices and unobserved demand shocks, we utilize the excise tax as a supply shifter. In particular, we employ the excise tax \((\tau^e)\) and the excise*low-income interaction \((\tau^e \times LI)\) as instruments for the posted price \((pp)\) and the posted price*low-income interaction \((pp \times LI)\). This identification strategy is valid under the same assumptions as the main specification, namely that cigarette tax changes are uncorrelated with unobserved shocks to cigarette demand. Tables 2.11 and 2.12 show that the results from the IV specification are similar to the specifications that omit pre-tax prices. Table 2.11 shows that both excise and sales tax changes are passed on slightly differently for high- and low-income consumers (e.g. retailers may decide how much to raise prices based on neighborhood income) but that these differences are quite small in magnitude, particularly for the sales-tax. Table 2.12 confirms that these differential pass-through rates do not drive our finding of increasing attentiveness by income.

2. Additional Robustness Checks

Appendix D investigates the robustness of our analysis to three additional concerns. First, our use of a two-part model for smoking demand may be biased by changes to the composition of the smoking population. To investigate this issue, we estimate smoking demand for the entire population with a linear regression and with a Tobit model censored at zero. Second, our results could reflect differences in the amount of time it takes high- and low-income consumers to learn about sales tax changes. Consequently, we include lagged tax rate values to determine whether the attentiveness gap fades over time. Finally, we try including state-specific time trends to account for the possibility that changes in a state’s tax rates are correlated with unobserved trends in that state’s smoking demand (such as anti-
smoking sentiment). As detailed in the Appendix, all three robustness checks are consistent with the results of the main analysis.

### III. Conclusion

Policymakers at all levels of government depend on commodity taxes to raise revenue, but such taxes are typically regressive, constituting a greater burden for low-income consumers. This paper has suggested a novel way for policymakers to lessen that regressivity: manipulating the fraction of the tax that is levied at the register as opposed to being included in a good’s posted price. In particular, we showed that levying a greater proportion of a commodity tax at the register shifts the tax’s burden away from attentive consumers. When low-income consumers pay more attention to register taxes than high-income consumers do, designing a tax in this way can lessen its regressivity. Conversely, when high-income consumers are the more attentive, imposing a commodity tax at the register will exacerbate its regressivity.

With this motivation in mind, we investigated whether high- and low-income consumers respond differently to register taxes on cigarettes. Exploiting state and time variation in tax rates, we found that low-income consumers reduce cigarette demand in response to both excise and sales taxes on cigarettes, whereas higher-income consumers only reduce cigarette demand in response to excise taxes. Although the empirical results do not allow us to definitively rule out alternative explanations, our findings are consistent with the hypothesis that attentiveness to cigarette register taxes declines by income. Hence, policymakers may be able to ease the financial burden of cigarette taxes on the poor by levying such taxes at the register instead of including them in cigarettes’ posted price.

How important are these welfare effects quantitatively? To provide a rough idea, recall from Part I that the welfare effect for attentive consumers of a revenue-neutral shift towards
register taxes stemmed from the effect of the shift on the after-tax price of \( x \),
\[
\frac{d(t_p + t_r)}{dt_r} \bigg|_{R} \]
From Equation (2.5), we can express the combined tax change in terms of estimable quantities,
\[
\frac{d(t_p + t_r)}{dt_r} \bigg|_{R} = \frac{\tau \epsilon A + \epsilon B \phi B}{1 - \tau (\epsilon A \phi A + \epsilon B \phi B)}.
\]
In our sample, the (weighted) average ratio of taxes to the after-tax price is \( \tau = 0.33 \). Determining the share of cigarettes consumed by attentive consumers (\( \phi_A \)) is complicated by the fact that our empirical procedure is designed to assess whether income differences in attentiveness exist, rather than identify exactly which consumers are attentive and which are not. In particular, our results suggest that the bottom income quartile of consumers are more attentive to cigarette register taxes than higher income consumers, but the evidence in Tables 2.7 and 2.8 is consistent with consumers in the second income quartile also falling into the attentive group. To be conservative, we compute the welfare effect assuming that only consumers below the 50th income percentile are in the attentive group; the magnitude of the effect increases when the attentive group is defined as consumers with income below the 25th percentile. From Table 2.2, we know that consumers above the 50th income percentile consume approximately 48 percent of cigarettes, so that \( \phi_A = 0.52 \) and \( \phi_B = 0.48 \). From Table 2.8, we compute the overall elasticity of cigarette demand with respect to the posted price to be \( \epsilon_A = 0.84 \) and \( \epsilon_B = 0.96 \). Using (2.5), these values imply
\[
\frac{d(t_p + t_r)}{dt_r} \bigg|_{R} = -0.21,
\]
so that a $1.00 increase in the cigarette register tax could accommodate a $1.21 reduction in the cigarette posted tax. For perspective, that revenue-neutral shift would free up approximately $77 a year for an attentive consumer who smokes a pack of cigarettes per day.\(^{66}\)

Three qualifications are important when interpreting our results. First, we have treated cigarettes as a standard consumption good, abstracting away from their addictive nature. However, the fact that cigarettes are addictive could alter the welfare implications of our results. For example, models along the lines suggested by Gruber and Koszegi (2004) or

\(^{65}\)For this approximation, we ignore the effect of the shift on the pre-tax price of \( x \). As Appendix B shows, that omission is justified when posted cigarette taxes are fully passed on to consumers, a condition consistent with the results in Table 2.10.

\(^{66}\)For comparison, defining the inattentive group threshold at the 25th income percentile implies that a revenue-neutral $1.00 increase in the register tax accommodates a $1.34 reduction in the posted tax, resulting in yearly savings of $123 for an attentive pack-a-day smoker.
Gruber and Mullainathan (2002) suggest that cigarette taxes can benefit consumer welfare when voters adopt such taxes as a method of exercising self-control; consequently, shifting a cigarette tax to the register could deprive some consumers of a valuable tool for self-discipline. At the other extreme, a rational-addiction model such as that presented in Becker, Grossman and Murphy (1994) would imply that cigarette consumption decisions are informed by consumers’ expectations concerning future prices; if such expectations are important, the demand equations employed here are misspecified.

Second, readers should be cautious about extrapolating our results to goods other than cigarettes. Although we have presented some evidence that attentiveness to cigarette register taxes declines by income, the cognitive cost model presented in Appendix C highlights the fact that this result can vary between goods. In particular, low-income consumers may well be less attentive to register taxes on goods that are relatively sensitive to income and that constitute a larger share of expenditures for high-income consumers. Moreover, Appendix A shows that in certain markets, shifting to a register tax has the potential to induce producers to raise a good’s pre-tax price. In particular, for goods characterized by elastic demand and inelastic supply, shifting to a register tax could actually worsen the burden of those taxes on all consumers, including the poor.

Finally, much of our analysis implicitly assumes that consumers’ attentiveness to register taxes is fixed. In reality, however, a revenue tax increase may drive some inattentive consumers to become attentive by increasing the utility loss from ignoring the tax (discussed in Appendix C). If one endogenizes the boundaries of the attentive and inattentive groups, a sufficiently large shift towards register taxes could necessitate a net increase in the combined tax rate if the register tax’s revenue advantage was more than offset by the reduction in revenue caused by some inattentive consumers becoming attentive. Similarly, our empirical specifications may be incomplete if high-income consumers’ attentiveness to cigarette register taxes depends on the size of the register tax already in place. Although our data lack the power to confirm that theoretical prediction convincingly, policymakers
should be cautious before adopting large shifts towards register taxes on the basis of results like ours.

Although our discussion has focused on taxes designed to raise revenue, the empirical findings presented here also speak to broader questions of tax design. For example, a number of public health advocates have suggested raising taxes on soft drinks as a way to combat population obesity, with some proponents calling for an expanded tax of any form on those products (Engelhard, Garson and Dorn (2009)) and others arguing that including the tax in the posted product price would be most effective (Brownell et al. (2009)). Our results suggest an important consideration is missing from this discussion, namely that taxes imposed at the register may affect the eating habits of high- and low-income consumers in different ways. Such issues deserve further investigation.
Table 2.1: Summary of Cigarette Tax Changes

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excise Tax</td>
<td>$0.29</td>
<td>$0.68</td>
<td>$0.50</td>
<td>$0.36</td>
<td>$1.43</td>
<td>$0.75</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>0.0%</td>
<td>7.5%</td>
<td>3.8%</td>
<td>0.0%</td>
<td>7.5%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Pre-Tax Price</td>
<td>$0.91</td>
<td>$1.28</td>
<td>$1.03</td>
<td>$1.84</td>
<td>$2.47</td>
<td>$2.20</td>
</tr>
<tr>
<td># State Changes</td>
<td>91</td>
<td>45</td>
<td>3</td>
<td>n/a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Excise taxes and prices are denoted in 2000 dollars.

Table 2.2: Average Cigarette Consumption by Income Quartile

<table>
<thead>
<tr>
<th>Smoking Rate (Extensive Margin, %)</th>
<th>Daily Consumption among Smokers (Intensive Margin, cigarettes)</th>
<th>Share of Total Consumption (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>Q1</td>
<td>Q1</td>
</tr>
<tr>
<td>1984</td>
<td>30.1</td>
<td>21.7</td>
</tr>
<tr>
<td>2000</td>
<td>29.2</td>
<td>20.1</td>
</tr>
<tr>
<td>All Years</td>
<td>29.1</td>
<td>17.4</td>
</tr>
<tr>
<td>Q4</td>
<td>23.7</td>
<td>11.9</td>
</tr>
<tr>
<td>All</td>
<td>27.8</td>
<td>17.6</td>
</tr>
</tbody>
</table>

# State Changes 91 45
# Federal Changes 3 n/a
Table 2.3: Effect of Taxes on Cigarette Demand - Extensive and Intensive Margins

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th>Intensive Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Excise Tax</td>
<td>-0.127***</td>
<td>-0.116***</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>-0.261*</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td>(0.140)</td>
<td>(0.100)</td>
</tr>
<tr>
<td>Income</td>
<td>-0.099***</td>
<td>-0.099***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.039***</td>
<td>-0.039***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>White</td>
<td>0.082***</td>
<td>0.082***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>H.S. Grad</td>
<td>-0.062***</td>
<td>-0.062***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>College Grad</td>
<td>-0.123***</td>
<td>-0.123***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Married</td>
<td>-0.067***</td>
<td>-0.067***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>0.086***</td>
<td>0.087***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Age</td>
<td>0.034***</td>
<td>0.035***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Age2</td>
<td>-0.001***</td>
<td>-0.001***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>Log Unemp. Rate</td>
<td>-0.028***</td>
<td>-0.028***</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Log State Income</td>
<td>-0.014</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Income Trend</td>
<td>-0.003**</td>
<td>-0.006***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Third- and fourth-order age polynomials are included in the regression but not displayed.
Outcome variables: probability of smoking (extensive) and log cigarette demand (intensive).
The F-stat is for the test of equality between the excise tax and the sales tax coefficients.
* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.4: Effect of Taxes on Cigarette Demand - Combined Effect

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excise Tax</strong></td>
<td>-3.749***</td>
<td>-3.508***</td>
<td>-3.577***</td>
</tr>
<tr>
<td></td>
<td>(0.481)</td>
<td>(0.479)</td>
<td>0.479</td>
</tr>
<tr>
<td><strong>Sales Tax</strong></td>
<td>-5.007**</td>
<td>-2.350</td>
<td>-2.366</td>
</tr>
<tr>
<td></td>
<td>(2.156)</td>
<td>(2.200)</td>
<td>2.199</td>
</tr>
</tbody>
</table>

| Economic Conditions | x            | x            |             |
| Income Trend        | x            | x            |             |
| **F-stat**          | 0.30         | 0.25         | 0.27        |
| **prob>F**          | 0.58         | 0.62         | 0.60        |
| **N**               | 1,288,031    | 1,288,031    | 1,288,031   |

Standard errors clustered at the state level in parentheses.

All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.

Outcome variable: cigarette demand in levels.

The F-stat is for the test of equality between the excise tax and the sales tax coefficients.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.5: Effect of Taxes on Cigarette Demand by Income - Extensive and Intensive Margins

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th>Intensive Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Excise Tax</td>
<td>-0.152***</td>
<td>-0.141***</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>-0.152</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.114)</td>
</tr>
<tr>
<td>Excise*Low-income</td>
<td>0.099*</td>
<td>0.099*</td>
</tr>
<tr>
<td></td>
<td>(0.054)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>Sales*Low-income</td>
<td>-0.502**</td>
<td>-0.502**</td>
</tr>
<tr>
<td></td>
<td>(0.191)</td>
<td>(0.191)</td>
</tr>
<tr>
<td>Income</td>
<td>-0.125***</td>
<td>-0.125***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Income Trend</td>
<td>-0.002**</td>
<td>-0.002**</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td></td>
</tr>
</tbody>
</table>

Economic Conditions  x  x  x  x  x
Income Trend         x  x  x  x
F-stat               9.91 9.74 8.60 5.48 5.49 5.14
prob>F               0.00 0.00 0.01 0.02 0.02 0.03
N                    1,288,031 1,288,031 1,288,031 274,137 274,137 274,137

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive) and log cigarette demand (intensive).
The F-stat is associated with the test for equality between the excise*low-income and sales*low-income interaction coefficients.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.6: Effect of Taxes on Cigarette Demand by Income - Combined Effect

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excise Tax</strong></td>
<td>-4.420***</td>
<td>-4.184***</td>
<td>-3.846***</td>
</tr>
<tr>
<td></td>
<td>(0.496)</td>
<td>(0.494)</td>
<td>(0.494)</td>
</tr>
<tr>
<td><strong>Sales Tax</strong></td>
<td>-2.290</td>
<td>0.365</td>
<td>0.265</td>
</tr>
<tr>
<td></td>
<td>(2.196)</td>
<td>(2.242)</td>
<td>(2.239)</td>
</tr>
<tr>
<td><strong>Excise*Low-income</strong></td>
<td>2.543***</td>
<td>2.554***</td>
<td>1.047*</td>
</tr>
<tr>
<td></td>
<td>(0.544)</td>
<td>(0.544)</td>
<td>(0.546)</td>
</tr>
<tr>
<td><strong>Sales*Low-income</strong></td>
<td>-11.987***</td>
<td>-11.989***</td>
<td>-11.888***</td>
</tr>
<tr>
<td></td>
<td>(1.915)</td>
<td>(1.916)</td>
<td>(1.917)</td>
</tr>
<tr>
<td><strong>Income</strong></td>
<td>-2.375***</td>
<td>-2.375***</td>
<td>-1.565***</td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td>(0.079)</td>
<td>(0.147)</td>
</tr>
<tr>
<td><strong>Income Trend</strong></td>
<td></td>
<td></td>
<td>-0.094***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Economic Conditions</strong></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Income Trend</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F-stat</strong></td>
<td>45.45</td>
<td>45.52</td>
<td>36.17</td>
</tr>
<tr>
<td><strong>prob&gt;F</strong></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1,288,031</td>
<td>1,288,031</td>
<td>1,288,031</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects. Outcome variable: cigarette demand in levels.
The F-stat is associated with the test for equality between the excise*low-income and sales*low-income interaction coefficients.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.7: Effect of Taxes on Cigarette Demand by Income Quartile - Extensive and Intensive Margins

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th></th>
<th></th>
<th>Intensive Margin</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Excise Sales</td>
<td>F-stat</td>
<td>Excise Sales</td>
<td>F-stat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (Q1)</td>
<td>-0.081</td>
<td>-0.521***</td>
<td>-0.329***</td>
<td>-1.072*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.166)</td>
<td>(0.101)</td>
<td>(0.566)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excise*Q2</td>
<td>-0.104**</td>
<td>0.284*</td>
<td>6.09**</td>
<td>0.015</td>
<td>0.883</td>
<td>3.04*</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.153)</td>
<td>(0.02)</td>
<td>(0.137)</td>
<td>(0.536)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Excise*Q3</td>
<td>-0.050</td>
<td>0.574**</td>
<td>6.88***</td>
<td>-0.019</td>
<td>1.551*</td>
<td>4.19**</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.220)</td>
<td>(0.01)</td>
<td>(0.103)</td>
<td>(0.802)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Excise*Q4</td>
<td>0.002</td>
<td>0.628***</td>
<td>9.92***</td>
<td>-0.069</td>
<td>1.875**</td>
<td>5.56**</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.190)</td>
<td>(0.00)</td>
<td>(0.130)</td>
<td>(0.839)</td>
<td>(0.02)</td>
</tr>
</tbody>
</table>

N 1,288,031 1,288,031 1,288,031 274,137 274,137 274,137

Standard errors clustered at the state level in parentheses in columns 1, 2, 4, and 5.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive) and log cigarette demand (intensive).
Tax*income group interactions represent the difference between that income group’s sensitivity to the tax rate and the baseline group’s sensitivity to the tax rate.
The F-stats are associated with testing $p_{2,j} - p_{1,j} = 0$ for $j$ in {2, 3, 4}. Prob>F in parentheses in columns 3 and 6.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.8: Effect of Taxes on Cigarette Demand by Income Quartile - Combined Effect

<table>
<thead>
<tr>
<th></th>
<th>Excise</th>
<th>Sales</th>
<th>F-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Baseline (Q1)</td>
<td>-2.917***</td>
<td>-11.575***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.641)</td>
<td>(2.663)</td>
<td></td>
</tr>
<tr>
<td>Excise*Q2</td>
<td>-1.628**</td>
<td>7.055***</td>
<td>10.85***</td>
</tr>
<tr>
<td></td>
<td>(0.665)</td>
<td>(2.347)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Excise*Q3</td>
<td>-0.999</td>
<td>13.686***</td>
<td>33.50***</td>
</tr>
<tr>
<td></td>
<td>(0.643)</td>
<td>(2.260)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Excise*Q4</td>
<td>-0.279</td>
<td>14.810***</td>
<td>33.65***</td>
</tr>
<tr>
<td></td>
<td>(0.660)</td>
<td>(2.323)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>N</td>
<td>1,288,031</td>
<td>1,288,031</td>
<td>1,288,031</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses in columns 1 and 2.

All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.

Outcome variable: cigarette demand in levels.

Tax*income group interactions represent the difference between that income group’s sensitivity to the tax rate and the baseline group’s sensitivity to the tax rate.

The F-stats are associated with testing $\rho_{2,j} - \rho_{1,j} = 0$ for $j$ in $\{2,3,4\}$.

Prob>F in parentheses in column 3.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.9: Sales Tax Exemptions for Cigarettes

<table>
<thead>
<tr>
<th></th>
<th>Extensive</th>
<th>Intensive</th>
<th>Combined</th>
<th>Extensive</th>
<th>Intensive</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td><strong>Excise Tax</strong></td>
<td>-0.124***</td>
<td>-0.338***</td>
<td>-3.701***</td>
<td>-0.124***</td>
<td>-0.333***</td>
<td>-3.693***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.061)</td>
<td>(0.688)</td>
<td>(0.029)</td>
<td>(0.061)</td>
<td>(0.688)</td>
</tr>
<tr>
<td><strong>Excise*Low-income</strong></td>
<td>0.045</td>
<td>-0.016</td>
<td>0.794</td>
<td>0.047</td>
<td>-0.035</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.114)</td>
<td>(1.395)</td>
<td>(0.060)</td>
<td>(0.115)</td>
<td>(1.401)</td>
</tr>
<tr>
<td><strong>Sales Tax*Non-exempt</strong></td>
<td>-0.451*</td>
<td>0.334</td>
<td>-7.146</td>
<td>-0.453*</td>
<td>0.358</td>
<td>-7.151</td>
</tr>
<tr>
<td></td>
<td>(0.262)</td>
<td>(0.572)</td>
<td>(6.656)</td>
<td>(0.262)</td>
<td>(0.573)</td>
<td>(6.656)</td>
</tr>
<tr>
<td><strong>Sales<em>Low-income</em>Non-exempt</strong></td>
<td>-0.501**</td>
<td>-1.450*</td>
<td>-11.872**</td>
<td>-0.489*</td>
<td>-1.587**</td>
<td>-11.884**</td>
</tr>
<tr>
<td></td>
<td>(0.249)</td>
<td>(0.755)</td>
<td>(5.812)</td>
<td>(0.253)</td>
<td>(0.762)</td>
<td>(5.911)</td>
</tr>
<tr>
<td><strong>Sales Tax*Exempt</strong></td>
<td>-0.392</td>
<td>-1.398</td>
<td>-13.158</td>
<td>-0.281</td>
<td>-2.037</td>
<td>-10.516</td>
</tr>
<tr>
<td></td>
<td>(0.364)</td>
<td>(1.355)</td>
<td>(9.193)</td>
<td>(0.338)</td>
<td>(1.752)</td>
<td>(9.538)</td>
</tr>
<tr>
<td><strong>Sales<em>Low-income</em>Exempt</strong></td>
<td>0.001</td>
<td>-0.276</td>
<td>0.224</td>
<td>-0.420</td>
<td>1.553</td>
<td>-9.849</td>
</tr>
<tr>
<td></td>
<td>(0.346)</td>
<td>(0.962)</td>
<td>(7.000)</td>
<td>(0.473)</td>
<td>(2.347)</td>
<td>(13.450)</td>
</tr>
<tr>
<td><strong>Low-income</strong></td>
<td>-0.016</td>
<td>0.023</td>
<td>-0.201</td>
<td>-0.016</td>
<td>0.030</td>
<td>-0.202</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.048)</td>
<td>(0.448)</td>
<td>(0.018)</td>
<td>(0.048)</td>
<td>(0.456)</td>
</tr>
<tr>
<td><strong>Exempt</strong></td>
<td>2.982***</td>
<td>2.813</td>
<td>75.639***</td>
<td>3.092**</td>
<td>-5.541</td>
<td>49.926*</td>
</tr>
<tr>
<td></td>
<td>(1.360)</td>
<td>(3.048)</td>
<td>(28.828)</td>
<td>(1.226)</td>
<td>(3.932)</td>
<td>(27.631)</td>
</tr>
<tr>
<td><strong>Exempt*Low-income</strong></td>
<td>0.018</td>
<td>-0.002</td>
<td>0.718</td>
<td>0.018</td>
<td>0.002</td>
<td>0.718</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.122)</td>
<td>(0.815)</td>
<td>(0.033)</td>
<td>(0.122)</td>
<td>(0.815)</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics, state, year, and calendar month fixed effects, and linear time trends interacted with exempt and low-income. Columns 4-6 include linear time trends in exempt*low-income. Outcome variables: probability of smoking (extensive), log cigarette demand (intensive), and cigarette demand in levels (combined). The F-stat is for the test of equality between the sales*low-income interaction between states that exempt cigarettes from the sales tax and states that do not.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table 2.10: States that Apply Excise Tax to Other Tobacco Products

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th>Intensive Margin</th>
<th>Combined Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Excise Tax</strong></td>
<td>-0.081***</td>
<td>-0.309***</td>
<td>-2.980***</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.101)</td>
<td>(0.625)</td>
</tr>
<tr>
<td><strong>Sales Tax</strong></td>
<td>0.319*</td>
<td>1.060</td>
<td>6.686**</td>
</tr>
<tr>
<td></td>
<td>(0.163)</td>
<td>(0.805)</td>
<td>(2.737)</td>
</tr>
<tr>
<td><strong>Excise*Low-income</strong></td>
<td>0.066</td>
<td>0.101</td>
<td>1.419</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.110)</td>
<td>(1.471)</td>
</tr>
<tr>
<td><strong>Sales*Low-income</strong></td>
<td>-0.757***</td>
<td>-1.512**</td>
<td>-16.982***</td>
</tr>
<tr>
<td></td>
<td>(0.178)</td>
<td>(0.690)</td>
<td>(4.503)</td>
</tr>
<tr>
<td><strong>F-stat</strong></td>
<td>17.52</td>
<td>5.47</td>
<td>15.22</td>
</tr>
<tr>
<td><strong>prob&gt;F</strong></td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>904,206</td>
<td>185,740</td>
<td>904,206</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive), log cigarette demand (intensive), and cigarette demand in levels (combined).
The F-stat is associated with the test for equality between the excise-poor and sales-poor interaction coefficients.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th>Intensive Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Excise Tax</td>
<td>1.018***</td>
<td>-0.296***</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>0.822*</td>
<td>0.185</td>
</tr>
<tr>
<td></td>
<td>(0.466)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Excise*Low-income</td>
<td>-0.041***</td>
<td>2.284***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.096)</td>
</tr>
<tr>
<td>Sales*Low-income</td>
<td>0.009</td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>R-square</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>N</td>
<td>1,288,031</td>
<td>1,288,031</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.

All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.

(1) and (3): Dependent variable = excise tax-inclusive price
(2) and (4): Dependent variable = excise tax-inclusive price*low-income
* p < 0.10, ** p < 0.05, *** p < 0.01
Table 2.12: Instrumenting for Cigarette Prices with Excise Tax

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th></th>
<th>Intensive Margin</th>
<th></th>
<th>Combined Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
<td>IV</td>
<td>OLS</td>
</tr>
<tr>
<td>Posted Price</td>
<td>-0.051***</td>
<td>-0.124***</td>
<td>-0.093***</td>
<td>-0.306***</td>
<td>-1.301***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.067)</td>
<td>(0.250)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>-0.040</td>
<td>0.070</td>
<td>0.227</td>
<td>0.557</td>
<td>-0.477</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.117)</td>
<td>(0.348)</td>
<td>(0.384)</td>
<td>(2.739)</td>
</tr>
<tr>
<td>Posted Price*Low-income</td>
<td>0.019</td>
<td>0.026</td>
<td>0.016</td>
<td>0.003</td>
<td>0.372</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.025)</td>
<td>(0.040)</td>
<td>(0.050)</td>
<td>(0.399)</td>
</tr>
<tr>
<td>Sales*Low-income</td>
<td>-0.492***</td>
<td>-0.495***</td>
<td>-1.411**</td>
<td>-1.368**</td>
<td>-11.547***</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.182)</td>
<td>(0.679)</td>
<td>(0.654)</td>
<td>(4.431)</td>
</tr>
<tr>
<td>F-test</td>
<td>8.44</td>
<td>8.04</td>
<td>4.68</td>
<td>4.73</td>
<td>7.80</td>
</tr>
<tr>
<td>prob&gt;F</td>
<td>0.01</td>
<td>0.00</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>N</td>
<td>1,288,031</td>
<td>1,288,031</td>
<td>274,137</td>
<td>274,137</td>
<td>1,288,031</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive), log cigarette demand (intensive), and cigarette demand in levels (combined).
Price includes the excise tax.
The F-stat is for the test of equality between the price*low-income and sales*low-income interaction coefficients.
*p < 0.10, **p < 0.05, ***p < 0.01
Figure 2.1: Average Monthly Taxes, 1984-2000

(a) Sales Tax

(b) Excise Tax

Figure 2.2: Aggregate Cigarette Consumption
Appendix

A. Welfare Analysis under Endogenous Producer Prices

This Appendix expands the model developed in Part I to the setting in which firms adjust their prices in response to changes in the type of tax imposed. As before, the policy we consider is an increase in the register tax coupled with a reduction in the posted tax calibrated to keep government revenue unchanged. Like CLK, we assume that taxes on \( x \) are fully-salient for producers.

Let \( p_x \) denote the after-tax price of \( p \), \( p_x \equiv p + t_p + t_r \). The net effect of the shift on the after-tax price of \( x \) is given by

\[
\frac{dp_x}{dt_r} = \frac{\partial p}{\partial t_r} + \frac{\partial t_p}{\partial t_r} + 1
\]  

(2.15)

Applying the same approach as in Part I, it is straightforward to show that the welfare effects of the shift for the two types of agents are given by

\[
\frac{dV_A}{dt_r} = -U_y(x_A, y_A), x_A \left(1 + \frac{\partial p}{\partial t_r} + \frac{\partial t_p}{\partial t_r}\right)
\]  

(2.16)

and
\[
\left. \frac{dV_B}{dt_r} \right|_R = -U_y(x_B, y_B) x_B \left( 1 + \left. \frac{\partial p}{\partial t_r} \right|_R + \left. \frac{\partial t_p}{\partial t_r} \right|_R \right) + \left( \left. \frac{\partial p}{\partial t_r} \right|_R + \left. \frac{\partial t_p}{\partial t_r} \right|_R \right) \frac{\partial x_B}{\partial p} \mu 
\]  

(2.17)

Let \( s(p) \) denote the supply of \( x \) as a function of \( x \)'s pre-tax price \( (p) \), so that the price-elasticity of supply is given by \( \varepsilon^S \equiv \frac{\partial s(p)}{\partial p} \times s \). Moreover, supply and demand of \( x \) must be equal in equilibrium:

\[
s(p) \equiv x_A(p + t_p + t_r) + x_B(p + t_p)
\]  

(2.18)

Totally differentiating (2.18) along with the government’s revenue constraint yields:

\[
\left. \frac{\partial p}{\partial t_r} \right|_R = \frac{\varepsilon_B \phi_B}{\varepsilon + \varepsilon^S (1 - \tau \varepsilon)} 
\]  

(2.19)

and

\[
\left. \frac{\partial t_p}{\partial t_r} \right|_R = -\frac{1 - \gamma \tau \varepsilon_A \phi_A}{1 - \gamma \tau \varepsilon} 
\]  

(2.20)

where \( \varepsilon \equiv \varepsilon_A \phi_A + \varepsilon_B \phi_B \) and \( \gamma \equiv \frac{\varepsilon^S}{\varepsilon^S + \varepsilon} \).  \(^6\)

Equation (2.19) shows that for \( \gamma < 1 \), a revenue-neutral shift towards register taxes results in a higher pre-tax price for all consumers. Because some consumers are more sensitive to posted taxes than to register taxes, replacing the former with the latter allows producers to shift some of the tax’s incidence back on to consumers. In turn, the higher pre-tax price reduces demand for \( x \), necessitating a larger \( t_p \) than otherwise in order for the government to meets its revenue constraint. Consequently, the reduction in the combined tax rate accommodated by the shift is smaller than when producer prices are fixed.

To illustrate, suppose that the supply of \( x \) is completely inelastic, \( \varepsilon^S = 0 \). What are the effects of a revenue-neutral increase in register taxes in this setting? As always, the increase in the register tax accommodates a reduction in the posted tax. Because \( \varepsilon^S = 0 \),

\(^6\)Recall that \( \tau \equiv \frac{t_p + t_r}{p + t_p + t_r} \), \( \varepsilon_i \equiv -\frac{x_i}{\partial p} \frac{p + x_i + t_i}{x_i} \) and \( \phi_i \equiv \frac{\varepsilon_i}{x_i + x_i} \). Note that \( \varepsilon \) \( \varepsilon < 1 \) follows from our maintained assumption that \( \frac{\partial R}{\partial p} > 0 \).
producers had absorbed the entire incidence of the posted tax; as \( t_p \) is reduced, the pre-tax price rises one for one. If all consumers were inattentive, the story would end here; for a $1 increase in the register tax, the posted tax would fall by \( \frac{\partial t_p}{\partial t_r} \) and the pre-tax price would rise by \( \frac{\partial t_p}{\partial t_r} \). When some consumers are attentive, the pre-tax price of \( x \) will fall somewhat in response to the new register tax; but as long as some consumers ignore the tax, producers will not have to reduce the pre-tax price in the full amount of the register tax increase. Hence the net effect of the shift on the after-tax price will be positive.

From (2.16), it is clear that a shift towards register taxes benefits attentive consumers if and only if the net effect of the shift on \( x \)'s after-tax price is negative. By substituting (2.19) and (2.20) into (2.15), it follows that:

\[
\left. \frac{dV_A}{dt_r} \right|_R > 0 \iff \left. \frac{dp_x}{dt_r} \right|_R > 0 \iff \tau \varepsilon^S > 1
\]

Thus when \( \varepsilon^S \) is sufficiently small, shifting towards a register tax makes even the attentive consumers worse off.\(^{68}\)

Similarly, \( \tau \varepsilon^S > 1 \) is a necessary condition for inattentive consumers to benefit from a shift towards register taxes. When \( \tau \varepsilon^S \leq 1 \), (2.21) implies that \( \left. \frac{dp_x}{dt_r} \right|_R \geq 0 \), which in turn implies that the first term in (2.17) is non-positive. Also, from (2.19) and (2.20), one can show that \( \left. \frac{dp}{dt_r} \right|_R + \left. \frac{\partial p}{\partial t_r} \right|_R \leq 0 \), implying that the second term in (2.17) is non-positive as well. Thus when \( \tau \varepsilon^S < 1 \), shifting from register to posted taxes makes all consumers worse off.

Finally, even when the supply of a taxed good is too inelastic for the government to raise welfare by shifting towards register taxes, the government’s choice between posted and register taxes still has important effects on consumer welfare. In particular, when \( \tau \varepsilon^S < 1 \), the government can raise the welfare of all consumers through a revenue-neutral shift

---

\(^{68}\)Another way to understand this dynamic is to observe that \( B \)'s inattentiveness to register taxes impose two distinct externalities on \( A \). First, \( B \)'s inattentiveness benefits \( A \) because it reduces the tax rate (which is levied on both \( A \) and \( B \)) needed for the government to obtain a given amount of revenue. Second, \( B \)'s inattentiveness harms \( A \) vis-a-vis producers because it reduces the overall market sensitivity to higher prices for \( x \). When some consumers are inattentive, demand for \( x \) does not fall as much in response to a given price increase, and consequently, producers do not have to reduce the pre-tax price of \( x \) by as much in order to maintain demand. As \( \varepsilon^S \) shrinks, the second externality grows in importance, and for small enough \( \varepsilon^S \), the second externality will dominate the first.
towards posted taxes – the opposite of the policy considered in Part I. Mechanically, this result follows directly from (2.16), (2.17), and (2.21). In words, when supply of the taxed good is sufficiently inelastic, producers will have to absorb the majority of the incidence of the new posted tax. Although the combined tax rate on \( x \) will increase, that increase will be more than offset by the reduction in the pre-tax price. Thus by increasing the salience of the tax for inattentive consumers, the government can precipitate a reduction in the market clearing price faced by attentive consumers. Attentive consumers are better off because of the net reduction in the after-tax price and inattentive consumers benefit both from the lower pre-tax price and because the associated reduction in register taxes reduces the magnitude of their optimization error.\(^{69}\)

**B. Welfare Analysis Under Alternate Budget Adjustment Rules**

Part I assumed that inattentive consumers who misperceive the price of \( x \) satisfy their budget constraints by reducing expenditures on \( y \). This Appendix considers the robustness of our results to alternate rules for mapping infeasible intended consumption bundles into feasible final consumption bundles.

In addition to the rule that we employ, Chetty, Looney, and Kroft (2007) identify two other “intuitive” budget adjustment rules. First, consumers who misperceive the price of \( x \) may satisfy their budget constraints by reducing expenditures on \( x \) rather than \( y \). This rule represents the other end of the spectrum from the one that we employ, and would be appropriate if consumers purchased \( x \) after completing their purchases of all other goods. Under this rule, it is easy to show that:

\[
\frac{\partial x_B}{\partial t_r} = \frac{-x_B}{p + t_r + t_p} \tag{2.22}
\]

\[
\frac{\partial x_B}{\partial t_p} = -\left( \frac{\partial y_B}{\partial p} + x_B \right) \frac{1}{p + t_r + t_p} \tag{2.23}
\]

\(^{69}\)Of course, whether or not such a welfare transfer is socially desirable depends upon how one values the trade off between consumer welfare and producer surplus.
The second alternate budget adjustment considered by Chetty, Looney, and Kroft (2007) is for inattentive agents to reduce consumption of both $x$ and $y$ to make up the income lost to the register tax. Inattentive consumers ignore the register tax when making their consumption decisions, but recognize that their net-of-tax income is lower because of the tax. For example, consumers who purchase $x$ and $y$ repeatedly will eventually realize that they consistently have less money in their bank account than they had anticipated. Inattentive consumers whose behavior is described by this rule will fully account for the tax’s income effect but fail to account for the tax’s substitution effect. As a result, we have:

\[
\frac{\partial x_B}{\partial t_r} = -x_B \frac{\partial I}{\partial I} \tag{2.24}
\]

\[
\frac{\partial x_B}{\partial t_p} = \frac{\partial x_B}{\partial t_r} + \frac{\partial \tilde{x}_B}{\partial p} \tag{2.25}
\]

where $\frac{\partial \tilde{x}_B}{\partial p}$ represents Hicksian (compensated) demand.

As before, we consider the welfare effects of a revenue-neutral shift from posted to register taxes. Because the attentive agent optimizes correctly, the welfare effect for that agent is the same as before:

\[
\left. \frac{dV_A}{dt_r} \right|_R = -U_y (x_A, y_A) x_A \left(1 + \left. \frac{\partial t_p}{\partial t_r} \right|_R \right)
\]

Totally differentiating the government’s budget constraint yields an expression for the posted tax reduction associated with a revenue-neutral increase in the register tax:

\[
\left. \frac{\partial t_p}{\partial t_r} \right|_R = -\frac{x_A + x_B + (t_p + t_r) (\frac{\partial x_A}{\partial t_r} + \frac{\partial x_B}{\partial t_r})}{x_A + x_B + (t_p + t_r) (\frac{\partial x_A}{\partial t_p} + \frac{\partial x_B}{\partial t_p})}
\]

A little algebra reveals that the welfare effect of the shift is positive for attentive consumers if and only if $\frac{\partial x_B}{\partial t_r} > \frac{\partial x_B}{\partial t_p}$, that is, when inattentive consumers reduce their demand for the taxed good by a larger amount in response to a posted tax increase than in response to a register tax increase. Intuitively, this condition ensures that the new register tax will be
more effective at raising revenue than the old posted tax was. Consequently, the shift accommodates a reduction in the combined tax rate, thus generating a positive income effect. Using (2.22) - (2.25), it is easy to see that this condition is satisfied under the two alternate budget adjustment rules.70

The welfare analysis for inattentive consumers proceeds as in Part I. Under the first alternate rule,

\[
\frac{dV_B}{dt_r} = -(1 + \frac{\partial p}{\partial t_r}) \frac{\partial x_B}{\partial p} \left( U_x(x_B,y_B) \left( \frac{\partial y_B}{\partial p} \frac{\partial p}{\partial t_r} \right) + \frac{\partial y_B}{\partial t_r} \left( U_y(x_A,y_A)(p + t_r + t_p) - U_x(x_A,y_A) \right) \right)
\]

Like the result in Part I, the welfare effect for inattentive consumers is ambiguous under this rule. Shifting to a register tax accommodates a reduction in the combined tax rate, generating a positive welfare effect (captured by the first term). Unlike before, however, the magnitude of this effect depends on the marginal utility of \(x\) rather than \(y\) because providing the consumer with additional income reduces the amount that the consumer must reduce her consumption of \(x\) to satisfy the budget constraint. The second term represents the cost of optimization error. Like before, this cost is zero when there are no register taxes and grows in size as register taxes push inattentive consumers further from their optimal bundle.

Under the second alternate rule, the welfare effect of the shift for inattentive consumers is also similar to that found in Part I. Here the welfare effect is given by

\[
\frac{dV_B}{dt_r} = -(1 + \frac{\partial p}{\partial t_r}) \frac{\partial x_B}{\partial p} \left( U_y(x_B,y_B)x_B + \left( \frac{\partial y_B}{\partial p} \frac{\partial p}{\partial t_r} - x_B \frac{\partial x_B}{\partial p} \right) \left( U_x(x_B,y_B) - (p + t_r + t_p)U_y(x_B,y_B) \right) \right)
\]

Again, the first term represents a positive income effect and the second term represents a negative welfare effect stemming from optimization error, which grows in size as register taxes increase.

\[C. A \text{ Cognitive Cost Model of Heterogeneous Attentiveness}\]

How does attentiveness to register taxes vary by income? The model we develop in this Appendix does not make a uniform prediction for all goods, but rather highlights the factors

\[\text{70Because } y \text{ represents all goods other than } x, \text{ it is reasonable to assume that } \frac{\partial y_B}{\partial p} > 0.\]
that determine which income group will be more attentive for a particular good. We then consider those factors in the context of cigarettes to predict whether high- or low-income consumers are likely to be more attentive to cigarette register taxes.

Suppose all agents have the option of paying attention to register taxes, but that doing so carries with it some positive utility cost.\footnote{The cognitive cost model we use as our starting point follows the basic approach laid out in Chetty, Looney, Kroft (2007).} This "cognitive cost" could stem from the mental effort needed to remember and calculate a good’s tax-inclusive price or might simply represent the opportunity cost of time spent on that task.

Assume that agents’ final utility is additively separable between the cognitive cost and consumption so that we can write $W_i = U(x_i, y_i) - b_i c_i$ in which $b_i$ is a binary choice variable indicating whether agent $i$ pays the cognitive cost and $c_i$ is the magnitude of the cost for agent $i$. We assume that the cognitive cost is fixed for a given individual in that it does not depend on the register tax rate (it requires just as much effort to take a 6 cent register tax into account as a 7 cent one).

The timing of the model with cognitive costs proceeds as in Part I, except here we add an initial step in which agents choose whether or not they will take register taxes into account when deciding on their consumption of $x$. As before, all agents choose an intended consumption bundle $(\hat{x}, \hat{y})$ subject to their perceived budget constraint, which we can now express as $\hat{BC}: x_i (p + b_i t_r + t_p) + y_i \leq M_i$.

A few final pieces of notation will be helpful. Let $(x^*_i, y^*_i)$ denote the (optimal) bundle that $i$ would consume if she were to pay attention to the register tax and let $(\tilde{x}, \tilde{y})$ denote the (sub-optimal) bundle she would consume were she to ignore the register tax. Agents who fail to pay the cognitive cost misperceive the after-tax price of $x$ as being lower than it actually is; as a result, they over-spend on $x$ and under-spend on $y$. The net change in $i$’s utility from taking the tax into account is therefore given by

$$W (x^*_i, y^*_i, 1) - W (\tilde{x}_i, \tilde{y}_i, 0) = G_i - c_i$$
where \( G_i \equiv U(x_i^*, y_i^*) - U(\bar{x}, \bar{y}) \) represents the agent’s utility gain from consuming the optimal feasible bundle.

We assume that agents opt to pay the cognitive cost when doing so affords them greater utility: \( b_i = 1 \{ G_i - c_i \geq 0 \} \). Although a full-fledged comparison between the utility that would be achieved in the two scenarios would likely require more cognitive effort than simply taking the tax into account in the first place, it seems reasonable that the agents who decide to pay the cognitive cost tend to be the ones for whom doing so has the most benefit.\(^{72}\)

Under the assumption that utility is additively separable in \( x \) and \( y \), Chetty, Looney, and Kroft (2007) show that one can express \( G_i \) (the gain in consumption utility from taking the tax into account) as

\[
G_i = \frac{1}{2} t^2 \varepsilon_{x,p} x_i^* v'(y_i^*) \left( \frac{1}{p + t} + \mu_i \gamma_i \right)
\]

where \( U(x, y) = u(x) + v(y) \), \( \varepsilon_{x,p} \) is the elasticity (defined to be positive) of \( x_i^* \) with respect to its price, \( \mu_i \equiv \frac{x_i}{y_i} \) represents the optimal ratio of \( x \) to \( y \), and \( \gamma_i \) measures the curvature of \( v() \) at \( y_i^* \): \( \gamma_i \equiv \frac{-v''(y_i^*)}{v'(y_i^*)} y_i^* \).

CLK allow differences in the extent to which individuals take taxes into account by assuming heterogeneity in the cognitive costs that agents face \( (c_i) \), although they do not model the sources of that heterogeneity. Because our goal is to link differences in attentiveness to agents’ income, we allow \( G_i \) to vary over individuals while abstracting from individual heterogeneity in cognitive costs: \( c_i = \bar{c}. \)\(^{73}\) In particular, we focus on individual heterogeneity that arises from differences in agents’ income. For a fixed tax rate and price,

\(^{72}\)Another justification for this approach is that agents might make a one-time comparison between \( G_i \) and \( c_i \) to decide whether to pay the cognitive cost in future circumstance. A third possibility is that agents decide attentiveness tax by tax, rather than good by good (as assumed here). If so, low-income consumers may be particularly attentive to sales taxes because such taxes constitute a relatively high share of their expenditures.

\(^{73}\)In reality, cognitive costs may also be correlated with income. The correlation may be positive, if high earners are better at cognitive tasks of this sort, or negative, if high earners have a greater opportunity cost of time. The extension to either of these cases is straightforward.
we can write $G_i$ as a function of the agent’s income ($M_i$)

$$G(M_i) = \frac{1}{2} t^2 \varepsilon_{x,p}(M_i) \left\{ \frac{x_i^*(M_i)}{p + t} + \mu_i(M_i) \gamma_i(M_i) \right\} v'(y_i^*(M_i))$$

The question we are interested in is whether low- or high-income individuals are more likely to take register taxes into account. Because agents are alike apart from their incomes, the question at hand is whether $G(\cdot)$ is increasing or decreasing in $M_i$. Differentiating the above expression with respect to income yields:

$$\frac{\partial G_i}{\partial M_i} = \frac{1}{2} t^2 \left\{ \frac{\partial \varepsilon_{x,p}}{\partial M_i} x_i^{**(p)} A v'(y_i^*) + \frac{\partial A}{\partial M_i} \varepsilon_{x,p} x_i^* v'(y_i^*) + \frac{\partial v'(y_i^*)}{\partial M_i} \varepsilon_{x,p} x_i^* A \right\} + \frac{\partial x_i^*}{\partial M_i} \varepsilon_{x,p} A v'(y_i^*)$$

where $A = \frac{1}{p + t} + \mu_i(M_i) \gamma_i(M_i)$. Since $A$, $x_i^*$, $\varepsilon_{x,p}$ and $v'(y_i^*)$ are all positive, the key terms to sign are $\frac{\partial \varepsilon_{x,p}}{\partial M_i}$, $\frac{\partial A}{\partial M_i}$, $\frac{\partial x_i^*}{\partial M_i}$, and $\frac{\partial v'(y_i^*)}{\partial M_i}$.

First, consider $\frac{\partial v'(y_i^*)}{\partial M_i}$. We know that $\frac{\partial v'(y_i^*)}{\partial M_i} = v''(y_i^*) \frac{\partial v^*}{\partial M_i} < 0$ assuming concave utility and that $y$ is a normal good. Intuitively, when the marginal utility of income declines rapidly with wealth, consumers who have little income to begin with are made much worse off by accidentally over-spending on $x$.

Second, consider $\frac{\partial x_i^*}{\partial M_i}$. This term will be positive as long as $x$ is a normal good, but will be smaller in magnitude for goods for which consumption does not much change as income rises. In words, consumers who consume more will gain more from optimizing correctly simply because the consumption difference caused by the optimization error will be larger in magnitude. When demand for $x$ is relatively insensitive to income, contribution of this term will be small.

Next consider $\frac{\partial \varepsilon_{x,p}}{\partial M_i}$. Are high- or low-income consumers more price sensitive in their demand for $x$? In general, theory is ambiguous as to whether elasticities rise or fall with income (the sign depends upon the magnitude of the third derivative of the utility function with respect to $x$).

Finally, consider $\frac{\partial A}{\partial M_i} = \frac{\partial \mu_i}{\partial M_i} \gamma_i + \frac{\partial \gamma_i}{\partial M_i} \mu_i$. Let’s take the two pieces in turn. $\frac{\partial x_i^*}{\partial M_i}$ is clearly positive as long as $x$ is a normal good. $\frac{\partial \mu_i}{\partial M_i}$ refers to how the optimal ratio of $x$ to $y$ changes
with income. This term is zero when preferences are homothetic and negative for consumption goods that constitute a larger share of expenditures for poor consumers than for rich consumers. The second term, $\frac{\partial \gamma_i}{\partial M_i}$, captures change in the curvature of utility from wealth as income rises; it will be weakly negative when consumers exhibit constant or decreasing relative risk aversion.

We have highlighted the factors that determine whether attentiveness to a register tax is increasing or decreasing by income. What does the analysis imply for the case of cigarettes? Regardless of the good in question, low-income consumers suffer more from lost consumption of other goods when they accidentally overspend on the taxed good. The key determinants that vary between goods are $\frac{\partial x}{\partial M_i}$, $\frac{\partial \varepsilon_i}{\partial p}$, $\frac{\partial \mu}{\partial M_i}$, and $\frac{\partial \mu}{\partial M_i}$.

For the case of cigarettes, all three of these factors suggest that attentiveness to register taxes should decrease by income. The income elasticity of cigarettes is generally found to be quite small (or even negative), implying a low value for $\frac{\partial x}{\partial M_i}$. Similarly, on average, poor households spend a substantially larger fraction of their income on cigarettes compared to rich households (Chaloupka and Warner 2000), which implies that $\frac{\partial \varepsilon_i}{\partial p} > 0$. Finally, the sign of $\frac{\partial \mu}{\partial M_i}$ hinges on whether low- or high-income consumers are more sensitive to cigarette prices. The empirical literature on this question is mixed, with most studies concluding that low-income smokers are slightly more price sensitive and other studies finding the opposite. In our data, we find the differences in price-sensitivity between rich and poor smokers to be small, implying that $\frac{\partial \varepsilon_i}{\partial p}$ is small in magnitude.

As a whole, our model suggests that attentiveness to cigarette register taxes should decline by income. Low-income consumers suffer more when they over-spend on cigarettes because their marginal utility of wealth is greater than that of high-income consumers. Although the magnitude of the optimization error will in general be larger for high-income consumers (the difference between their intended and realized bundles is bigger), this factor is mitigated in the case of cigarettes by the fact that smoking demand is relatively insensitive to income and by the fact that low-income consumers spend a substantially higher fraction of
their income on cigarettes compared to high-income consumers.

D. Additional Robustness Checks

This Appendix investigates the sensitivity of our analysis to additional robustness checks.

1. Alternative Specifications

So far, we have followed the approach taken by much of the smoking literature by separately modeling the extensive and intensive margins of cigarette consumption. This approach has the advantage of providing information about the mechanism by which tax changes reduce cigarette demand, in particular whether higher prices reduce demand by motivating smokers to quit or cut back. However, a drawback of this approach is that the intensive margin results may be biased by changes to the composition of the smoking population.\footnote{For example, suppose that smokers’ demand for cigarettes were completely insensitive to price changes, but that light smokers quit when the price became too high. In such a world, a tax increase would appear to raise the intensity of smoking demand on the intensive margin merely by raising the fraction of heavy smokers in the smoking population.}

As a robustness check, we estimate smoking demand using a linear regression and a Tobit model censored at zero. The dependent variable in these regressions is the number of cigarettes smoked per day, with the variable assigned a value of zero when the individual in question is not a smoker. Because the entire population of respondents is used, these approaches avoid the problem that tax rate changes affect selection into the smoking population. The flip side of the coin is that these models do not allow variables to differ in how they affect smoking demand on the intensive and extensive margins. Moreover, the Tobit specification relies on the normality of the unobservables and the linear functional form is probably unrealistic for an application in which so many of the observations have a dependent variable equal to zero. The results of the linear and Tobit specifications are presented in Appendix Table 1 and are consistent with the results from the two-part model.
used in the rest of the paper.

2. Delayed Responses to Tax Changes

So far we have assumed that smoking demand depends only upon current cigarette taxes, but it could be that tax changes affect consumer behavior with a lag. For example, higher prices might motivate smokers to quit, but the quitting process itself could take several months. Alternatively, it could be that consumers take some time to learn about sales tax changes, only gradually incorporating them into their behavior. If these lags were different for high- and low-income consumers, it could provide an alternative explanation for our results. To investigate this issue, we examine the sensitivity of our results to using various lags of the tax rates instead of the current rate.

\[
y_{ismt} = \alpha + \beta_1 \tau_{sm,t-k}^e + \beta_2 \tau_{sm,t-k}^s + \rho_1 \tau_{sm,t-k} LI_{ismt} + \rho_2 \tau_{sm,t-k} LI_{ismt} + \eta LI_{ismt} + \\
\gamma x_{smt} + \delta z_{ismt} + \mu_s + \lambda_t + \pi_m + \epsilon_{ismt}
\]  

(2.26)

where \( k \) is three, six, or twelve months. The results are reported in Appendix Table 2 and suggest that our results are not being driven by differences in the time it takes high- and low-income consumers to respond to cigarette tax changes.

3. State Specific Trends

Although including state fixed-effects accounts for unobserved factors that affect the levels of smoking demand by state, it could be that changes in a state’s tax rates are correlated with trends in that state’s cigarette demand, such as anti-smoking sentiment. To reduce the influence of any such omitted third factors, we add state-specific year trends to the

\[^7^5\text{For example, high-income consumers may be better able to afford top of the line smoking-cessation products.}\]
Although our tax rate data is probably largely free of measurement error, including state trends could still cause substantial attenuation bias in the current context. Suppose that smoking demand depends upon a function of current and past tax rates, $x_t = x(a(L)x_t)$, where $a(L)$ is some lag polynomial. The situation here is analogous to the standard measurement error problem: although the original tax variable $x_t$ may be highly correlated with the “true” tax variable $a(L)x_t$, the new tax measure after including state trends may only be weakly correlated with the “true” tax rate, causing an attenuation bias.

$y_{ismt} = \alpha + \beta_1 \tau_{smt} + \beta_2 \eta_{ismt} + \rho_1 \tau_{smt} L_{ismt} + \rho_2 \tau_{smt} p_{ismt} + \gamma x_{smt} + \delta \bar{z}_{ismt} + \mu_s + \lambda_t + \xi_s * t + \pi_m + \epsilon_{ismt}$

(2.27)
Table 2.14: Timing (Appendix Table 2)

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th></th>
<th>Intensive Margin</th>
<th></th>
<th>Combined Effect</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td></td>
<td>3 Month</td>
<td>6 Month</td>
<td>12 Month</td>
<td>3 Month</td>
<td>6 Month</td>
<td>12 Month</td>
</tr>
<tr>
<td>Excise Tax</td>
<td>-0.116***</td>
<td>-0.104***</td>
<td>-0.112***</td>
<td>-0.349***</td>
<td>-0.332***</td>
<td>-0.259***</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.067)</td>
<td>(0.061)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>-0.057</td>
<td>-0.156</td>
<td>-0.133</td>
<td>0.363</td>
<td>0.417</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.105)</td>
<td>(0.085)</td>
<td>(0.352)</td>
<td>(0.308)</td>
<td>(0.386)</td>
</tr>
<tr>
<td>Excise*LI</td>
<td>0.046</td>
<td>0.034</td>
<td>0.038</td>
<td>0.048</td>
<td>0.042</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.065)</td>
<td>(0.069)</td>
<td>(0.113)</td>
<td>(0.113)</td>
<td>(0.126)</td>
</tr>
<tr>
<td>Sales*LI</td>
<td>-0.460**</td>
<td>-0.470***</td>
<td>-0.471***</td>
<td>-1.420**</td>
<td>-1.427*</td>
<td>-1.360*</td>
</tr>
<tr>
<td></td>
<td>(0.179)</td>
<td>(0.167)</td>
<td>(0.154)</td>
<td>(0.672)</td>
<td>(0.721)</td>
<td>(0.717)</td>
</tr>
<tr>
<td>F-stat</td>
<td>7.63</td>
<td>8.51</td>
<td>10.86</td>
<td>5.51</td>
<td>4.55</td>
<td>3.94</td>
</tr>
<tr>
<td>prob&gt;F</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>N</td>
<td>1,285,448</td>
<td>1,282,765</td>
<td>1,277,073</td>
<td>273,406</td>
<td>272,636</td>
<td>271,088</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive), log cigarette demand (intensive), and cigarette demand in levels (combined).
The F-stat is associated with the test for equality between the excise*low-income and sales*low-income interaction coefficients.

*p < 0.10, **p < 0.05, ***p < 0.01
Table 2.15: State Trends (Appendix Table 3)

<table>
<thead>
<tr>
<th></th>
<th>Extensive Margin</th>
<th>Intensive Margin</th>
<th>Combined Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Excise Tax</td>
<td>-0.017</td>
<td>-0.348***</td>
<td>-1.357**</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.095)</td>
<td>(0.685)</td>
</tr>
<tr>
<td>Sales Tax</td>
<td>0.122</td>
<td>0.342</td>
<td>3.436</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.425)</td>
<td>(3.336)</td>
</tr>
<tr>
<td>Excise*Low-income</td>
<td>0.060</td>
<td>0.019</td>
<td>1.078</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.112)</td>
<td>(1.379)</td>
</tr>
<tr>
<td>Sales*Low-income</td>
<td>-0.501***</td>
<td>-1.378**</td>
<td>-11.814***</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.665)</td>
<td>(4.380)</td>
</tr>
<tr>
<td>F-stat</td>
<td>8.62</td>
<td>5.09</td>
<td>8.98</td>
</tr>
<tr>
<td>prob&gt;F</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>N</td>
<td>1,288,031</td>
<td>274,137</td>
<td>1,288,031</td>
</tr>
</tbody>
</table>

Standard errors clustered at the state level in parentheses.
All specifications include individual demographic characteristics and state, year, and calendar month fixed effects.
Outcome variables: probability of smoking (extensive), log cigarette demand (intensive), and cigarette demand in levels (combined).
The F-stat is associated with the test for equality between the excise*low-income and sales*low-income interaction coefficients.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Chapter 3

Preference Identification Under Inconsistent Choice

(with Daniel Reck)

Abstract

In many settings seemingly arbitrary features of a decision may nonetheless affect choice. We develop an empirical framework to recover ordinal preference information from binary choice data characterized by framing effects. Plausible restrictions of varying strength permit either partial- or point-identification of preferences for frame-insensitive decision-makers. Recovering population preference requires understanding the empirical relationship between decision-makers’ preferences and their susceptibility to framing. We develop tools for studying this relationship and illustrate them with data on automatic enrollment into pension plans. The results suggest 60 percent of default-sensitive employees prefer enrollment but that defaults of non-enrollment may be optimal for younger, lower-income employees.

Introduction

Suppose a government wishes to regulate how internet companies collect and analyze their customers’ personal data. Some customers prefer that companies collect their data, perhaps to help predict what content they will enjoy; others prefer that their data be kept private. A key question in designing such policies is whether privacy controls should be opt-in – so that customers must give a company permission before it can use their data – or opt-out – so that a company can use its customers’ data unless a customer tells it otherwise. Although both policy designs let customers control the use of their personal data, empirical research suggests the framing of the decision affects which option customers will choose (e.g. Johnson, Bellman and Lohse, 2002). For example, it may be that under the opt-in policy 40 percent of customers allow a company to use their data but that under the opt-out policy 70 percent do so. If the government’s only goal is to maximize customers’ welfare, how should it decide whether privacy controls should be opt-in or opt-out?
Answering questions like this through the lens of economic theory is complicated by the difficulty in identifying preferences\textsuperscript{77} when behavior varies according to seemingly arbitrary features of the choice environment – such as which option is the default, the order in which options are presented, or which features of the options are made salient. When choices vary according to factors that are unrelated to decision-makers’ preferences, equating those choices with preferences – as the revealed preferences approach does – is problematic. The question of how to conduct welfare analysis in such settings is at the heart of important controversies in behavioral economics. In particular, for a benevolent government to design choices in ways that maximize decision-makers’ well-being, it must first have some means of identifying the preferences of those decision-makers whose behavior will be affected.

Most prior work takes one of two approaches to addressing the problem of preference recovery under inconsistent choice. First, researchers may utilize a positive model of behavior that fully specifies the mapping from a decision-maker’s preferences to her (potentially sub-optimal) behavior (e.g. Rubinstein and Salant, 2012; Benkert and Netzer, 2014; Carroll et al., 2009). Such approaches yield important insights but in many cases the resulting welfare conclusions are sensitive to the modeler’s choice between competing positive models that are difficult to distinguish observationally (Bernheim, 2009; De Clippel and Rozen, 2014). An alternative approach is to restrict preference inferences to the subset of observed choice situations in which a given decision-maker chooses consistently (Bernheim and Rangel, 2009). However, in practice individual decision-makers are typically observed making only a single choice, which makes it difficult to detect which choices are consistent. Worse, this approach yield no information on the preferences of those decision-makers who exhibit systematic choice reversals – the very group whose preferences are

\textsuperscript{77}By preferences, we mean the relative consistency of the available options with a decision-maker’s objectives, which we take to be the normatively-relevant measure of the decision-maker’s welfare, as in Rubinstein and Salant (2012). Preferences are not defined according to a decision-maker’s observed choices as under the standard revealed preferences approach; doing so would assume away the question we address by (tautologically) ruling out the possibility of choice reversals. For discussions of this issue, see Basu (2003) and Sen (1973; 1977; 1993).
most relevant for determining the optimal policy (a claim we formalize in the Online Appendix). Further “refinements” can provide a path forward if the researcher can observe choices in a setting in which all decision-makers are known to choose optimally Chetty, Looney and Kroft (e.g., 2009), but in many applications, such as those in which behavior is sensitive to defaults or ordering effects, there is little reason to believe that all of the decision-makers choose optimally in any of the observed choice situations.\textsuperscript{78}

In this paper we develop a new framework for preference identification when decision-makers exhibit systematic choice reversals. We focus primarily on binary choice settings in which the option chosen by some decision-makers varies according to a preference-irrelevant feature of the choice environment and provide conditions under which one can identify the distribution of preferences in the population of observed decision-makers as well as within various subgroups of that population. We do so by combining a relaxation of the revealed preferences approach suggested by Bernheim and Rangel (2009) with the potential outcomes framework commonly used in the literature on causal identification (Angrist, Imbens and Rubin, 1996).\textsuperscript{79} This innovation permits us to recover preference information with weaker datasets – those in which each individual is observed making only one decision – and without assuming that all decision-makers choose optimally in any one observed choice situation. In addition, we provide a range of conditions under which one can recover the preferences of those decision-makers who exhibit choice reversals. Our primary focus is on preference identification, but we also introduce new tools for investigating heterogeneity in which decision-makers commit choice reversals. Although our approach requires additional structure relative to Bernheim and Rangel (2009), it retains an

\textsuperscript{78}Another possibility is to turn from actual to hypothetical choice data designed to elicit preference parameters (Barsky et al., 1997), or more radically, away from preference-based measures of well-being altogether (e.g., Benjamin et al., 2012; Kahneman, Wakker and Sarin, 1997). While useful, such approaches are subject to criticisms of their own: for example, any survey-based method is potentially subject to numerous framing effects (e.g., Schwarz and Clore, 1987; Deaton, 2012) and approaches divorced from individual preferences may fail to capture normatively-important components of welfare (Loewenstein, 1999). A useful discussion of these and other issues related to behavioral preference recovery is provided in Beshears et al. (2008).

\textsuperscript{79}Unlike other applications of the potential outcomes framework, our goal is not to identify the causal effects of one variable on another, but rather to remove variation in observed choices due to framing effects, thus isolating the variation due to preferences.
important “reduced-form” flavor that allows researchers to draw conclusions about welfare in the face of uncertainty about the exact positive model that generates behavior.\textsuperscript{80}

To begin, we follow Salant and Rubinstein (2008) and Bernheim and Rangel (2009) by modeling decisions in terms of menus and frames (preference-irrelevant features of the choice environment that affect behavior). For the most part, we restrict our focus to binary menus and binary frames. Examples of frames might include: (1) which option is presented as the default; (2) the order in which options are displayed; (3) whether the consequence of selecting an option is presented as a loss or a gain; (4) whether the menu of options includes an irrelevant alternative; (5) the point in time at which a decision is made; or (6) whether various consequences of the available options are made salient.

When decision-makers choose consistently across frames, we assume those choices reflect their preferences. We refer to this assumption as the \textit{consistency principle}. We also allow decision-makers to choose inconsistently across frames, but initially we limit our analysis to settings in which the frame pulls the choices of all decision-makers in a uniform direction, an assumption we label \textit{frame monotonicity}. Crucially, our approach does not require that an outside observer be able to identify ex ante which decision-makers are optimizing, nor that an individual decision-maker be observed in multiple frames. Instead, we exploit the fact that frame monotonicity and the consistency principle jointly imply that a decision-maker who chooses “against the frame” prefers the option that she chooses. This insight, along with a statistical assumption concerning the assignment of decision-makers to frames, allows us to recover preferences among the consistent decision-makers.\textsuperscript{81} It also allows us to identify the distribution of observable characteristics among the consistent and inconsistent decision-makers, even though members of those groups cannot be individually

\textsuperscript{80}To illustrate this point, the Online Appendix demonstrates the compatibility of our approach with a range of structural models for why decision-makers might exhibit sensitivity to default effects.

\textsuperscript{81}An alternative interpretation of the empirical evidence concerning “choice reversals” is to conclude that inconsistent decision-makers simply lack normatively relevant preferences in the first place. For someone who takes that view as a starting point, the contribution of our paper is that it provides a method for backing out the (normatively relevant) preferences of the consistent decision-makers from the aggregate observed choice data.
identified. Without frame monotonicity, the preferences of the consistent decision-makers are partially-identified, and we provide the corresponding bounds.

We next consider what can be learned about population preferences from the distribution of preferences among consistent choosers. We begin by showing how our assumptions permit partial identification of population preferences. Full identification of population preferences requires understanding the empirical relationship between decision-makers’ preferences and their consistency. Intuitively, the first step of our analysis yields preference information for a subset of the population (the consistent decision-makers). By understanding the relationship between decision-makers’ likelihood of selecting into that sub-population and their likelihood of having a particular preference, we can extrapolate from the preferences of the consistent decision-makers to the full population. When consistency is uncorrelated with decision-makers’ preferences – a condition we refer to as \textit{decision quality independence} – population preferences may be recovered by extrapolating the preferences of the consistent decision-makers directly to those whose behavior varies by the frame. When decision quality independence does not hold, we provide two approaches for shedding light on the empirical relationship between decision-makers’ consistency and their preferences.

The first approach is to adjust for observable differences between the consistent and inconsistent groups and then extrapolate from one to the other. If decision quality independence holds \textit{conditional} on these observable characteristics, one can recover population preferences by separately estimating the preferences of each demographic group and then re-weighting those estimates based on the distribution of observables among the inconsistent decision-makers. For example, it may be that the poor are more likely to be influenced by the frame than the rich, and also that the poor have different preferences than the rich; but that conditional on income, decision-makers’ susceptibility to the frame is uncorrelated with their preferences. As in other empirical contexts, the plausibility of this matching-on-observables approach depends on what information about decision-makers can be ob-
served.

The second approach we develop exploits variation in the decision-making environment related to individuals’ susceptibility to the frame. A decision quality instrument has the following two properties: (1) it monotonically affects decision-makers’ propensity to choose consistently, and (2) conditional on whether an individual is consistent, it does not affect choice. For example, a decision quality instrument could take the form of the time pressure under which a decision must be made: decision-makers faced with greater time pressure may be more likely to choose according to the frame, but time pressure is unlikely to affect which option they actually prefer. Other potential decision quality instruments could include the intensity of the frames, the costs of comparing the available options, or the presence or absence of various drains on the decision-maker’s “cognitive load.” Variation in a decision quality instrument sheds light on the correlation between decision-makers’ preferences and their propensity to choose consistently, the key unknown needed for recovering population preferences. In particular, this approach permits observers to recover the distribution of preferences for the set of decision-makers whose susceptibility to the frame is affected by the decision quality instrument.82 We then describe two extrapolation techniques for using this information to estimate population preferences.

We illustrate our framework using data on automatic enrollment and participation in employer-sponsored pension plans drawn from Madrian and Shea (2001). We document a strong positive relationship between employees’ consistency across default regimes (opt-in versus opt-out) and their preferences for enrollment in the pension plan. The results suggest that although most of the default-sensitive employees in the firm we study prefer enrollment, a sizable minority (32 percent) do not. Preferences for non-enrollment are disproportionately concentrated among younger and lower-income employees, suggesting there may be value to customizing default options based on employee characteristics.

The paper proceeds as follows: Section 1 sets up the model. Section 2 provides point-

---

82 As discussed below, these assumptions and results parallel the identification of a Local Average Treatment Effect using instrumental variables (Imbens and Angrist, 1994).
and partial-identification conditions for the preferences of the consistent choosers. Section 3 characterizes the problem of recovering preferences for the full population. Section 3.1 derives bounds for population preferences; Section 3.2 develops the matching-on-observables approach; and Section 3.3 develops the decision quality instrument approach. Section 4 illustrates our approach using data on defaults and enrollment into employer-provided pension plans. The Online Appendix contains proofs of propositions (A); motivates our parameters of interest with a simple model of optimal frame design (B); considers the relationship between our framework and alternative structural models of default effects (C); generalizes the framework to settings with non-binary frames and non-binary menus (D); and derives standard errors for finite-sample inference (E).

1. Setup

This section introduces the notation and assumptions employed throughout the paper. We presume to observe individual choice data from a population of density 1, with individuals denoted by $i$. The observed decisions are binary, $y_i \in \{0,1\}$, and each decision-maker is observed under exactly one of two possible frames, denoted $d_0$ and $d_1$.\footnote{Our definition of a frame is based on Salant and Rubinstein (2008) and Bernheim and Rangel (2009). In settings where the frame is multi-dimensional, such as variation in which option is the default and the order in which the options are presented, we can apply this framework using the two most extreme frames — those that make decision-makers most likely and least likely to choose $y = 1$, respectively — as $d_1$ and $d_0$. See Online Appendix D for generalizations beyond the two-option, two-frame setting.} Let $y_{1i}$ and $y_{0i}$ denote what $i$ would choose under $d_1$ and $d_0$, respectively. Population moments are given by $Y_1 \equiv E[y_{1i}|d_i = d_1]$ and $Y_0 \equiv E[y_{0i}|d_i = d_0]$. To illustrate the notation using the using the privacy example from the introduction, $y$ could indicate whether an individual allows a company to use her data, so that $d_1$ would indicate the opt-out regime, and $d_0$ would indicate the opt-in regime; the population moments are $Y_1 = 0.7$ and $Y_0 = 0.4$. We assume throughout that population moments such as these are directly observable, setting aside issues of finite-sample statistical inference.
Decision-makers have ordinal, asymmetric preferences over the available options, denoted by $y_i^* \in \{0, 1\}$. Implicit in this notation is the following assumption:

**A1 (Frame Separability)** For all individuals, $y_i^*$ does not depend on $d$.

Frame separability limits which features of the decision-making environment are treated as a frame. Features of a decision that affect choice but that are relevant to decision-makers’ preferences over the available options are not frames.\(^{84}\) Importantly, frame separability does not require decision-makers to be irrational; a decision-making feature that imposed a transaction cost for selecting one of the options would constitute a frame, as long as it did not also affect decision-makers’ ultimate preference for ending up with one option or the other.\(^{85}\)

Decision-makers may either choose consistently or choose in a way that is sensitive to the frame. We denote consistency by $c_i \equiv 1\{y_{1i} = y_{0i}\}$. We assume throughout that the fraction of consistent decision-makers is strictly positive, $E[c_i] > 0$. When a decision-maker chooses consistently, we assume that her choice reflects her preferences.

**A2 (The Consistency Principle)** For all individuals, $c_i = 1 \implies y_i = y_i^*$.

In the privacy settings example described above, the consistency principle implies that a customer who would choose to keep his data private under both the opt-in and opt-out frame would in fact prefer that his data be kept private. The consistency principle relaxes the instrumental rationality assumption relied on by neoclassical revealed preference analysis, in that the choices made by inconsistent decision-makers need not reveal their preferences. It fails when decision-makers suffer from biases that cause them to make the same mistake under every frame in which they are observed.

---

\(^{84}\)For example, if a decision-maker chooses hot chocolate from \{hot chocolate, ice cream\} under one frame and ice cream from \{hot chocolate, ice cream\} under the other frame, there would be no apparent deviation from rationality if the frame indicated whether the season was winter or summer. This assumption is explicit in Salant and Rubinstein (2008) and implicit in Bernheim and Rangel (2009), who require it for determining when two potentially conflicting choice situations differ in terms of the frame or in terms of the available menu items. In this sense, frame separability is the property that distinguishes variation in frames from variation in menu items.

\(^{85}\)Put differently, $y_i^*$ indicates which option a rational individual would select in the absence of transaction costs.
Because each decision-maker is observed under only one frame, consistency is not directly observable from the data. If consistency were directly observable, Assumptions A1 and A2 alone would permit the identification of consistent decision-makers’ preferences, as in Bernheim and Rangel (2009). The following two assumptions permit us to recover this information under weaker data requirements.

\[ \text{A3 (Unconfoundedness)} (y_{1i}, y_{0i}) \perp d_i. \]

Unconfoundedness is a statistical assumption about the process by which decision-makers are assigned to frames. It ensures that differences in observed choices under different frames is due to the effect of the frames rather than to differences in the decision-makers assigned to them. Unconfoundedness is guaranteed when frames are randomly assigned.

\[ \text{A4 (Frame Monotonicity)} \text{ For all individuals, } y_{1i} \geq y_{0i}. \]

Frame monotonicity requires that when a frame affects choice, it does so in the same direction for each affected decision-maker. In the privacy settings example described above, frame monotonicity fails if some customers choose to allow access to their data if and only if doing so is not the default. Most of our discussion assumes frame monotonicity but we also derive partial identification results for settings in which the assumption fails.

2. Identifying Consistent Preferences

We initially focus on consistent decision-makers, those whose behavior is not affected by the frame. Recovering the preferences of this group would be trivial if decision-makers were observed under each frame; in that case an observer could identify which decision-makers were consistent and, using the consistency principle, which options the consistent decision-makers preferred. However, many real-world datasets do not have this property, and even when they do, the order in which decision-makers are exposed to frames may itself affect behavior (LeBoeuf and Shafir, 2003). The following proposition provides conditions for the identification of consistent decision-makers’ preferences when each decision-maker
is observed under a single frame:

**Proposition 1**  Let \( Y_c \equiv \frac{Y_0}{y_0 + 1 - y_1} \).

(1.1) Under A1 - A4, \( E[y_1^+ | c_i = 1] = Y_c \).

(1.2) Under A1 - A3, \( Y_c \geq \frac{1}{2} \iff E[y_1^+ | c_i = 1] \geq Y_c \).

**Proof**  By construction, \( (y_1i, y_0i) \in \{(1, 1), (0, 0), (1, 0), (0, 1)\} \). Frame monotonicity rules out \( (y_1i, y_0i) = (0, 1) \). Therefore we know that \( y_0i = 1 \iff (y_1i, y_0i) = (1, 1) \), and by the consistency principle, \( (y_1i, y_0i) = (1, 1) \implies y_i^+ = 1 \). Thus, \( E[y_0i] = Pr(y_i^+ = 1; c_i = 1) \). By the same logic, frame monotonicity and the consistency principle imply that \( E[1 - y_1i] = Pr(y_i^+ = 0; c_i = 1) \). Then by definition, \( P(c_i = 1) = E[y_0i] + E[1 - y_1i] \), and by the definition of conditional probability, \( E[y_1^+ | c_i = 1] = \frac{E[y_0i]}{E[y_0i] + E[1 - y_1i]} \). By unconfoundedness, \( Y_1 = E[y_1i] \) and \( Y_0 = E[y_0i] \); substituting these into the previous expression yields 1.1.

The proofs of 1.2 and of all further results are contained in Online Appendix A. □

Proposition 1.1 follows from the insight that, under frame monotonicity, only consistent decision-makers choose against the frame (i.e. choose \( y = 1 \) when confronted with \( d_0 \) or choose \( y = 0 \) when confronted with \( d_1 \)). Unconfoundedness guarantees that the assignment of individuals to frames is uncorrelated with preferences or consistency, which means that we can treat the set of decision-makers choosing against the frame as a representative sample of all consistent choosers. Finally, the consistency principle ensures that the observed choices of this group reveal the preferences of the corresponding decision-makers. As a result, the denominator of \( Y_c \) measures the fraction of decision-makers that are consistent and the numerator measures the subset of that group with \( y_i^+ = 1 \).

Proposition 1.2 provides a partial identification result that is robust to failures of frame monotonicity. **Frame-defiers** are those inconsistent decision-makers who select \( y_i = 1 \iff d = d_0 \). The presence of frame-defiers means that some decision-makers who are inconsistent will be misclassified as consistent. The misclassified group will contain all frame-
defiers, plus an equal and offsetting number of inconsistent decision-makers who are not frame-defiers. Because it contains the two types of inconsistent decision-makers in equal proportions, this group chooses \( y_i = 1 \) with probability \( \frac{1}{2} \) under both frames. Misclassifying this group of decision-makers as consistent therefore biases \( Y_c \) toward \( \frac{1}{2} \).

Table 3.1 illustrates this result for the hypothetical data on online privacy choices described in the introduction, where \( y_i = 1 \) if the customer allows a company to use her data. Under frame monotonicity, we can conclude that 70 percent of individuals are consistent across default regimes and that 57 percent of those customers prefer allowing the company to use their data. Without frame monotonicity, we may only conclude that at least 57 percent of the consistent customers prefer allowing the company to use their data.

<table>
<thead>
<tr>
<th>Table 3.1: Aggregate Choices by Frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction choosing ( y = 1 )</td>
</tr>
<tr>
<td>Fraction choosing ( y = 0 )</td>
</tr>
</tbody>
</table>

| Fraction consistent, \( E[c_i] \), under A1-A4 | \( 0.4 + 0.3 = 0.70 \) |
| Consistent preferences, \( E[y_i^* | c_i = 1] \), under A1-A4 | \( \frac{0.4}{0.7} = 0.57 \) |
| Bounds on \( E[y_i^* | c_i = 1] \), under A1-A3 | [0.57, 1] |

The preference information recovered by Proposition 1 is important for several reasons. First, if one’s philosophical starting point is that inconsistent decision-makers lack normatively relevant preferences (e.g., Fischhoff, 1991), Proposition 1 is the end-point of the analysis. The value of the result is that it provides a method of isolating the normatively-relevant parameter (the consistent decision-makers’ preferences) from the noise induced by the frames. Second, when population preferences are known – what Bernheim and Rangel (2009) refer to as a “refinement” – Proposition 1 can be used in conjunction with that information to recover the preferences of the inconsistent decision-makers.\(^{86}\) Such information

\[ E[y_i^* | c_i = 0] = \frac{E[y_i^*] - E[y_i^* | c_i = 1] E[c_i]}{1 - E[c_i]} \]
is often valuable because optimal policy may turn on the preferences of the inconsistent decision-makers (see Online Appendix B), but observing aggregate population preferences under a refinement does not provide the preferences for that subgroup. Finally, the preferences of the consistent decision-makers may be used to recover the preferences of the remainder of the population by accounting for selection into the consistent sub-population, which is the task we undertake in the remaining sections.

3. Identifying Population Preferences

The remainder of the paper focuses on using the preferences of the consistent decision-makers to gain information about the preferences of the inconsistent decision-makers or of the full population. The basic challenge to doing so is overcoming a potential selection bias: when selection into the consistent sub-population is not random, characteristics of the consistent decision-makers may be correlated with the preferences of that group. In many ways, this challenge parallels the well-known problem of selection into treatment that has been studied in the program evaluation literature. However, an important difference is that in the typical sample selection context, the researcher can identify which units have been selected into the relevant sample. In contrast, whether a particular decision-maker is consistent is unobservable when each decision-maker is observed under a single frame.

To clarify the selection challenge, note that we can write

$$E[y^*_i] = E[y^*_i | c_i = 1] - \frac{cov(y^*_i, c_i)}{E[c_i]}$$

(3.1)

where the equation follows from the identity $cov(y^*_i, c_i) = E[y^*_i c_i] - E[y^*_i]E[c_i]$ and the fact that $E[y^*_i c_i] = P(y^*_i = c_i = 1) = E[y^*_i | c_i = 1]E[c_i]$. Equation (3.1) highlights that recovering population preferences from consistent sub-group preferences requires accounting for the

---

87 For example, assessing the effect of a job training program on wages may be biased if the program induces some individuals to become employed when they would not have been employed otherwise (e.g., Lee, 2009). In that context, the researcher can observe whether a given individual has wage data and hence whether he or she has been selected into the sample of employed workers.
correlation between preferences and consistency (the other parameters in the equation are identified under A1-A4). Moreover, the covariance is a sufficient statistic for identifying population preferences despite uncertainty about the underlying behavioral model; that is, for the purposes of identifying $E[y^*_i]$, the behavioral model only matters to the extent that it shapes $\text{cov}(y^*_i, c_i)$. Note that in the special case in which the covariance term is zero – a condition we refer to as decision quality independence – the preferences of the consistent decision-makers will be representative of the full population.\footnote{Decision quality independence is analogous to the familiar “missing at random” assumption in the sample selection literature.}

3.1 Partial Identification

Absent information on the relationship between preferences and consistency, the distribution of preferences in the population may be partially identified in the spirit of Manski (1989).

**Proposition 2**

(2.1) Under A1-A4, $E[y^*_i] \in [Y_0, Y_1]$.

(2.2) Under A1 -A3, $\max \{Y_0 - (1 - Y_1), 0\} \leq E[y^*] \leq \min \{Y_0 + Y_1, 1\}$.

The partial identification result in (2.1) is quite intuitive: with frame monotonicity, the fraction preferring some option lies between the fraction choosing that option under the two frames. When the fraction of inconsistent decision-makers is large, the bounds will be relatively uninformative.

Without frame monotonicity, we obtain weaker, one-directional bounds for population preferences. The result follows from noting that $E[y^*_i]$ depends on three parameters: $E[y^*_i | c_i = 0]$, $E[y^*_i | c_i = 1]$, and $E[c_i]$. Although $E[y^*_i | c_i = 0]$ is unobservable, the other two parameters can be inferred from the data given information on the prevalence of frame-defiers. Knowing that $E[y^*_i | c_i = 1] \in [0, 1]$ constrains the prevalence of frame defiers, which
then yields bounds on the value of $E[y^*_i]$. The further $Y_0$ is from $1 - Y_1$, the more informative the bounds will be.\(^{89}\) Note that when frame monotonicity fails, it is possible that $Y_0$ and $Y_1$ lie on the same side of $\frac{1}{2}$ but that $E[y^*_i]$ lies on the other side of $\frac{1}{2}$.

Using the hypothetical data from Table 3.1, we would conclude under frame monotonicity that the fraction of the population preferring that their personal data be used is between 40 and 70 percent. Without monotonicity, we can only conclude that this fraction is greater than 10 percent.

To summarize the results thus far, the degree to which $E[y^*_i]$ and $E[y^*_i | c_i = 1]$ can be identified from the data depends on the strength of the researcher’s assumptions about the behavioral model. When only A1-A3 are imposed, the data permit partial identification of $E[y^*_i]$ and $E[y^*_i | c_i = 1]$, where the bounds on the former are wider than those on the latter. Adding frame monotonicity permits $E[y^*_i | c_i = 1]$ to be point identified and narrows the bounds on $E[y^*_i]$. Finally, imposing decision quality independence permits point identification of $E[y^*_i]$ as well. The remaining two sections provide alternative identification conditions for $E[y^*_i]$ which rely on frame monotonicity but not decision quality independence.

### 3.2 Matching on Observables

In some settings, the relationship between preferences and consistency may be driven by factors that are observable to the researcher, such as income, education, age, or prior experience with the decision at hand. For example, it could be that more educated customers are less likely to prefer that companies use their personal data and more likely to choose consistently across default regimes, but that conditional on education, preferences and consistency are independent. This section develops identification strategies for these settings.

\(^{89}\)When $Y_0 = 1 - Y_1$ exactly, the bounds are entirely uninformative because the data do not constrain the fraction of frame-defiers and, as a result, we cannot rule out $E[c_i] = 0$. Consequently, when $Y_0 = 1 - Y_1$, any $E[y^*_i] \in [0, 1]$ is feasible.
of these observables permits us to relax the unconfoundedness assumption:

**A3’ (Conditional Unconfoundedness).** For all observable characteristics \( w \), \((y_{1i}, y_{0i}) \perp d_i \mid w_i = w \).

Using the observable characteristics to extrapolate from the preferences of consistent decision-makers requires the following assumption:

**A5 (Conditional Decision Quality Independence)** For all individuals and all observable characteristics \( w \), \( \text{cov}(y^*_i, c_i \mid w_i = w) = 0 \).

Conditional decision quality independence requires that consistent and inconsistent decision-makers with the same observable characteristics have the same distribution of preferences. As with any matching-on-observables approach, the plausibility of this assumption will depend on the detail and quality of the observable characteristics as well as the underlying positive model of behavior. We examine this question in more detail in Online Appendix C. In general, A5 is more likely to hold when variation in consistency is driven by heterogeneity in the cost of optimizing or in the tendency to employ a psychological heuristic (C.1.2), rather than intensity in preferences over the available options (C.2).

The identification strategy we propose in this section is: first, to estimate the preferences of consistent decision-makers with given observable characteristics; second, to extrapolate preferences from consistent to inconsistent decision-makers with the same observable characteristics; and third, to use weighted combinations based on the distribution of observable characteristics to recover preferences in the full population or the sub-population of inconsistent decision-makers.

An important barrier to employing this familiar approach in our context is that we cannot directly observe consistency. The following lemma shows that the distribution of characteristics among the consistent and inconsistent decision-makers are nonetheless identified.\(^{90}\)

\(^{90}\)In this sense, Lemma 1 is analogous to Abadie (2003), who shows how to identify the aggregate observable characteristics of compliers with respect to an instrument when individual compliers cannot be identified.
Lemma 1 Let $Y_j(w) = E[y_{j|i} | d_i = d_j, w_i = w]$ for $j = 0, 1$, $q_w = \frac{Y_0(w) + 1 - Y_1(w)}{E_w[Y_0(w) + 1 - Y_1(w)],}$ and $s_w = \frac{Y_1(w) - Y_0(w)}{E_w[Y_1(w) - Y_0(w)]}$. Under A1, A3', and A4:

(L1.1) For any $w$, $p(w_i = w | c_i = 1) = q_w p(w_i = w)$

(L1.2) For any $w$, $p(w_i = w | c_i = 0) = s_w p(w_i = w)$.

Apart from its role as a step in the matching estimator, Lemma 1 is useful in its own right. Information on the observable correlates of consistency is important for researchers investigating the mechanisms by which frames affect decision-making and for policymakers designing interventions aimed at particular sub-groups of the population.91 Exploiting Lemma 1 along with conditional decision quality independence, the following proposition formalizes the matching-on-observables identification strategy described above:

Proposition 3 Let $Y_c(w) = \frac{Y_0(w)}{Y_0(w) + 1 - Y_1(w)}$. Under Assumptions A1, A2, A3', A4, and A5:

(3.1) $E[y^*_i] = E_w[Y_c(w)]$

(3.2) $E[y^*_i | c_i = 0] = E_w[s_w Y_c(w)]^{92}$

Table 2 illustrates the identification approach for our privacy controls example, reporting hypothetical data conditioned upon whether individuals have a high school education. The population moments in the third column match the moments in Table 3.1. The conditioning reveals that high-school-educated individuals are more likely to be consistent and the

Continuing with the analogy, Proposition 3 is related to Angrist and Fernandez-Val (2010), who exploit information on the distribution of observables to extrapolate an estimated treatment effect from one subset of a population to another.

91 For example, Thaler and Sunstein (2008) advocate designing frames in ways that offset other decision-making biases but for this approach to work, it must be that those decision-makers that are subject to the bias being targeted are also the ones who are sensitive to the frame being set. Lemma 1 helps answer this question by allowing the researcher to determine which decision-makers are likely to be sensitive to a given frame. Lemma 1 is also valuable for assessing which types of decision-makers are “more rational” when the researcher is unable to observe repeated decisions by individual decision-makers, as required for the approach developed in Choi et al. (2014).

92 Replacing assumption A3 with A3’ in (1.1) and (1.2) implies that $E[c_i] = E_w[Y_0(w) + 1 - Y_1(w)]$, and $E[y^* | c_i = 1] = E_w[q_w Y_c(w)]$. The results in Proposition 3 make use of this revised estimator for $E[c_i]$. Even under random frame assignment, the revised estimator for $E[c_i]$ will be preferable for applications of Proposition 3 in finite sample, due to possible spurious correlation between observables and frame assignment. In particular, using the revised estimator ensures the weights implied by (3.1) will sum to one.
consistent choosers among them are more likely to prefer that the company use their personal data \((y_i = 1)\). Under conditional decision quality independence, we conclude that 44 percent of inconsistent decision-makers, and 53 percent of the population prefer that their personal data be used. Under \textit{unconditional} decision quality independence, both these fractions would be 57 percent and we would over-estimate the share preferring that their personal data be used.

<table>
<thead>
<tr>
<th>Table 3.2: Average Choices by Frame and High School Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Fraction choosing (y = 1) under (d_1, Y_1(w))</td>
</tr>
<tr>
<td>Fraction choosing (y = 1) under (d_0, Y_0(w))</td>
</tr>
<tr>
<td>Fraction of population, (p(w))</td>
</tr>
<tr>
<td>Fraction consistent, (E[c_i</td>
</tr>
<tr>
<td>Fraction of consistent population, (p(w</td>
</tr>
<tr>
<td>Fraction of inconsistent population, (p(w</td>
</tr>
<tr>
<td>Consistent preferences, (E[y^*</td>
</tr>
<tr>
<td>Inconsistent preferences, (E[y^*</td>
</tr>
<tr>
<td>Population preferences, (E[y^*_i])</td>
</tr>
</tbody>
</table>

### 3.3 Decision Quality Instruments

In many settings, selection into the consistent subpopulation will be driven by factors that are unobservable to the researcher. In this section we develop an identification strategy that does not rely upon conditional or unconditional decision quality independence. Specifically, we introduce the notion of a \textit{decision quality instrument}, which exploits variation in the decision-making environment that affects decision-makers' consistency but that is orthogonal to their preferences. The methods developed here share many similarities with traditional instrumental variables in empirical economics but, as we explore below, differ from traditional instrumental variable analysis in important ways.

Let \(z\) denote a decision quality instrument with two values, \(z \in \{z_h, z_l\}\). Individual choices now depend on \(d\) and \(z\); we denote them by by \(y_{ijk}\), where \(j \in \{0, 1\}\) indexes the frame and \(k \in \{h, l\}\) indexes the instrument. Consistency is defined at each value of
the instrument and denoted by \( c_{ik} = 1\{y_{i1k} = y_{i0k}\} \). We denote the fraction of decision-makers observed choosing \( y \) under a given \((d, z)\) combination by \( Y_{jk} \equiv E[y_{i,jk} | d = d_j, z = z_k] \).

The following assumptions establish which variation constitutes a valid decision quality instrument:

**A3** (Unconfoundedness of \( d \) and \( z \)) \((y_{i1h}, y_{i0h}, y_{i1l}, y_{i0l}) \perp (d_i, z_i)\)

**A6** (Decision quality exclusivity) For all individuals, \( y^*_i \) does not depend on \( z \).

**A7** (Decision quality monotonicity) For all individuals, \( c_{ih} \geq c_{il} \) with \( E[c_{ih} - c_{il}] > 0 \).

Assumption A3" modifies the unconfoundedness assumption, which now requires that both \( d \) and \( z \) be uncorrelated with confounding factors. Assumption A6 requires that variation in the decision-making environment induced by \( z \) is irrelevant from the perspective of decision-makers’ preferences; it ensures that \( z \) affects behavior by altering consistency, not by changing which option decision-makers prefer.\(^{93}\) Assumption A7 requires that the effect of \( z \) on consistency is weakly monotonic for all decision-makers and strictly monotonic for some.

Variation in \( z \) might arise from natural experiments or be induced by researchers. For example, suppose that some decision-makers were randomly assigned to a treatment group aimed at manipulating their “cognitive load” – such as by memorizing a 10 digit number – prior to making the decision being studied. Such experimental designs could plausibly manipulate decision-makers’ susceptibility to a frame in ways that are unrelated to their preferences. Other examples of decision quality instruments might include the time pressure for making a decision, the cost of obtaining or processing information about the available choices, the opportunity cost of cognitive resources at the time of decision-making, or the intensity of the frame (e.g., the degree to which one alternative is more salient than another).

\(^{93}\)Like A1, A6 does not rule out variation in \( z \) affecting welfare by altering the transaction costs associated with choosing against the frame. Indeed, exogenous variation in such costs are excellent candidates for decision quality instruments. See Online Appendix C.
3.3.1 Identifying Sometimes-Consistent Preferences

This section develops a reduced-form approach to recover the preferences of those decision-makers whose consistency is affected by a decision quality instrument.

**Proposition 4.** Assume that A1, A2, and A4 hold at each fixed value of z, and assume A3", A6, and A7. Then

\[ E[y_i^*|c_{ih} > c_{il}] = \frac{Y_{0h} - Y_{0l}}{Y_{1l} - Y_{0l} - (Y_{1h} - Y_{0h})} \]

**Discussion of Proposition 4** Proposition 4 is best understood by analogy to the identification of a local average treatment effect (Imbens and Angrist, 1994). The monotonicity assumption (A7) permits us to divide the population into three groups of decision-makers: the always-consistent \((c_{ih} = c_{il} = 1)\), the sometimes-consistent \((c_{ih} = 1; c_{il} = 0)\), and the never-consistent \((c_{ih} = c_{il} = 0)\). The denominator of the expression in Proposition 4 measures the decrease in the size of the inconsistent sub-group as we move from \(z_l\) to \(z_h\), which identifies the size of the sometimes-consistent group. The expression in the numerator measures the change in the fraction choosing \(y = 1\) under \(d_0\) as \(z\) changes, which identifies the fraction of decision-makers who are sometimes-consistent and prefer \(y = 1\).

Several other parallels to the instrumental variables literature are apparent. First, one can use Proposition 4 to motivate over-identification tests of decision quality independence along the lines of Wu (1973); Hausman (1978). However, such a test requires \(E[y^{*}|c_{ih} > c_{il}] = E[y^{*}]\), which may fail depending on the nature of selection into consistency. We explore less restrictive alternatives below. Additionally, the types of variation that will satisfy assumptions A6 and A7 depends on the underlying model of behavior that generates framing effects, reflecting a familiar interplay between structural reasoning and instrumental variables. We discuss this issue further in Online Appendix C. Finally, Proposition 4 may be extended beyond binary decision quality instruments, by applying Proposition 4 to each pair-wise combination of values of \(z\) or, when \(z\) is continuous, by
adapting the methods of Yitzhaki (1989) (see also Heckman and Vytlacil, 2007).94

Table 3.3 illustrates this identification approach for the hypothetical data on privacy controls. We now suppose that the process for adjusting privacy settings may be either onerous (customers are required to navigate through several web pages) or streamlined (customers may adjust privacy settings with a single click). Aggregate choices under the onerous design correspond to the population moments reported in Table I. Customers are less susceptible to default effects when the process is streamlined. We can back out the fraction of consistent choosers and the aggregate preferences of the consistent choosers at either $z_l$ or $z_h$ using Proposition 1, as before. Note that the fraction of consistent customers who prefer that the company use their personal data is lower under $z_h$ than $z_l$, because the variation in $z$ affects $Y_1$ more than $Y_0$. Applying Proposition 4 implies that of the 20 percent of the population of customers who are sometimes-consistent, only 25 percent prefer that the company use their data, a share substantially below that of the consistent choosers at either $z_h$ or $z_l$.

Table 3.3: Average Choices by Frame and Difficulty of Changing Privacy Settings

<table>
<thead>
<tr>
<th></th>
<th>Onerous ($z_l$)</th>
<th>Streamlined ($z_h$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction choosing $y = 1$ under $d_1$</td>
<td>0.70</td>
<td>0.55</td>
</tr>
<tr>
<td>Fraction choosing $y = 1$ under $d_0$</td>
<td>0.40</td>
<td>0.45</td>
</tr>
<tr>
<td>Fraction consistent, $E[c_i]$</td>
<td>0.70</td>
<td>0.90</td>
</tr>
<tr>
<td>Consistent preferences, $E[y^*_i</td>
<td>c_{ik} = 1]$</td>
<td>0.57</td>
</tr>
<tr>
<td>Sometimes consistent preferences, $E[y^*_i</td>
<td>c_{ih} &gt; c_{il}]$</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The next two sections develop identification conditions for population and inconsistent decision-maker preferences that utilize variation in $z$. On its own, Proposition 4 does not identify these parameters; rather, by shedding light on the covariance between preferences and consistency, it allows us to extrapolate preference information from consistent

94 Another use for Proposition 4 is motivated by the optimal policy problem facing governments that must choose which $z$ value to implement, for example a regulator deciding how streamlined privacy controls should be. Online Appendix B shows the solution to this problem trades off the cost of selecting a $z$ that induces greater consistency against the welfare gain from doing so. The latter depends on the preferences of the decision-makers who choose consistently at one candidate $z$ but not in another, which Proposition 4 can be used to estimate.
decision-makers to other groups in the population.

### 3.3.2 Structural Extrapolation with Decision Quality Instruments

This section develops a latent variable model of the relationship between decision-makers’ consistency and their preferences, assuming a bivariate normal distribution for the idiosyncratic terms. Suppose that consistency for individual $i$ is determined by

$$ P_i = \overline{P} + \theta z_i + \varepsilon_i \quad (3.2) $$

$$ c_i = 1 \iff P_i > 0 \quad (3.3) $$

where $P_i$ is a latent variable reflecting idiosyncratic variation $\varepsilon_i$ and the effect of a binary decision quality instrument $z_i \in \{0, 1\}$. Note that consistency depends on $i$’s choice under both frames, so $P_i$ does not depend on the frame to which $i$ is assigned. Note also that decision quality monotonicity (A7) is satisfied provided $\theta \neq 0$.

Next, suppose the distribution of preferences can also be described with a latent variable model:

$$ M_i = \overline{M} + \nu_i \quad (3.4) $$

$$ y^*_i = 1 \iff M_i > 0 \quad (3.5) $$

where latent variable $M_i$ simply reflects idiosyncratic variation in preferences, $\nu_i$. Frame separability (A1) is satisfied because $M_i$ does not depend on $d$, and decision quality exclusivity (A6) is satisfied because $M_i$ does not depend on $z_i$. Unconfoundedness (A3”) is satisfied provided that $\varepsilon_i$ and $\nu_i$ are independent of $z_i$ and $d_i$.

Assume that $\varepsilon_i$ and $\nu_i$ are characterized by a bivariate standard normal distribution:

$$ \begin{pmatrix} \varepsilon_i \\ \nu_i \end{pmatrix} \sim N \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (3.6) $$

where $\rho \in (-1, 1)$ is the correlation between the error terms and where the normalization is without loss of generality. Note that decision quality independence is satisfied if and only
if $\rho = 0$. We close the model with the consistency principle (A2) and frame monotonicity (A4). Together, these assumptions allow us to evaluate the probability of observing a given choice for a given $(d, z)$:

\[
\forall i, \forall k = 0, 1, y_{i0k} = 1 \iff \epsilon_i > -\bar{P} - \theta z_k; \nu_i > -\bar{M}; \quad (3.7)
\]

\[
\forall i, \forall k = 0, 1, y_{i1k} = 0 \iff \epsilon_i > -\bar{P} - \theta z_i; \nu_i < -\bar{M}; \quad (3.8)
\]

Equations (3.7) and (3.8) can be combined with (3.6) to identify the parameters of the model ($\bar{P}, \bar{M}, \theta, \text{and } \rho$). We can then recover ordinal preferences by integrating the underlying distribution:

\[
E[y^*_i] = \Phi(\bar{M}), \quad \Phi() \text{ is the standard normal cumulative density function, and } E[y^*_i | c_{ik} = 0] = \frac{1}{1 - E[y^*_i | c_{ik} = 0]} \int_{-\infty}^{\bar{P} - \theta z_k} \int_{-\bar{M}}^{\bar{M}} \phi^{BVSN}(\epsilon, \nu; \rho) \partial \nu \partial \epsilon, \quad \phi^{BVSN}(a, b; \rho) \text{ is the bivariate standard normal density with correlation coefficient } \rho \text{ evaluated at } (a, b).
\]

The structural model described above resembles the classic bivariate normal model of Heckman (1979). Variation in the decision quality instrument induces variation in consistency without affecting preferences; this guarantees the relationship between consistency and preferences is identified without relying on functional form alone (Puhani, 2000).95

Applying the model to the data from Table III yields an estimated correlation coefficient of $\rho = 0.54$; the positive estimate suggests decision-makers with a high propensity to choose consistently (so that they are consistent at $z_t$) are more likely to prefer $y$ than those with a low propensity to choose consistently. The estimated parameters imply $E[y^*_i] = 0.46$. The population average is below both $E[y^*_i | c_{il} = 1]$ and $E[y^*_i | c_{ih} = 1]$ because it incorporates the preferences of the decision-makers with the very lowest propensity to choose consistently.

95 With a binary decision decision quality instrument, the model is just-identified. Additional values of $z$ permit maximum likelihood estimation of the model’s parameters.
3.3.3 Semi-Parametric Extrapolation with Decision Quality Instruments

This section develops an extrapolation approach that avoids relying on functional form assumptions to identify population preferences. In particular, we model the preferences of the consistent decision-makers at a given value of the decision quality instrument as a flexible polynomial in the fraction of decision-makers who are consistent at that value of the decision quality instrument.\[96\]

Suppose the decision quality instrument is observed taking on \(N+1\) values, indexed \(z_0, z_1, \ldots, z_N\), and drawn from a continuous ordered set of values, \([z, \bar{z}] \subseteq \mathbb{R}\) such that \(E[c_{i\bar{z}}] = 0\) and \(E[c_{iz}] = 1\). In addition, suppose that decision quality monotonicity holds with respect to any two values of \(z\):

\[A7'\] For all individuals and all \(z, z' \in [z, \bar{z}]\) such that \(z > z'\), \(c_{iz} \geq c_{iz'}\) and \(E[c_{iz} - c_{iz'}] > 0\)

For each individual, let \(z^*_i < \bar{z}\) denote the value of \(z\) at which she begins to choose consistently, i.e., \(z \geq z^*_i \implies c_{iz} = 1\). Assumption \(A7'\) implies that \(z^*_i\) is unique. Denoting the CDF of \(z^*_i\) by \(F(.)\) and the PDF by \(f(.)\), we have \(E[c_{iz}] = F(z)\). In addition, note that the second part of \(A7'\) guarantees \(f(z) > 0\) for all \(z \in [z, \bar{z}]\), so that \(F(.)\) is strictly increasing with a well-defined inverse function over \(E[c_{iz}] \in [0, 1]\), which we denote \(F^{-1}(E[c_{iz}])\).

Finally, let \(g(z) = E[y^*_i|z^*_i = z]\) denote the preferences of the marginally consistent decision-makers at a given \(z\).\[97\] To guarantee the validity of the Taylor Series approximation that underpins the following result, it will be convenient to assume that both \(F(z)\) and \(g(z)\) are infinitely differentiable with respect to \(z\).

**Proposition 6** Under \(A1, A2, A6, \text{ and } A7'\), for any degree \(D \in \mathbb{N}\), there are constants \(a_0 \ldots a_D\) such that

---

96This approach shares some similarity to the literature on non-parametric identification of marginal treatment effects from local average treatment effects (Heckman and Vytlacil, 2005). An important difference is that the techniques in that literature utilize instrumental variables that drive the propensity to participate in the treatment over a range from 0 to 1. But in our context, if we were able to observe decisions made under a decision-quality state that induced everyone to choose consistently, we could simply look at the preferences revealed in that state to recover the preferences for the population.

97Although by definition \(z \in \mathbb{R}\), the value of \(z\) itself may be unobservable to the researcher.
Proposition 6 implies that the preferences of the consistent decision-makers at a particular value of the decision quality instrument can be approximated by a polynomial function in the fraction of decision-makers who choose consistently at that value of the instrument. Because $A7'$ guarantees a one-to-one mapping between $z$ and $E[c_{iz}]$, we can write the preferences of the marginally consistent decision-makers as a function of the fraction of decision-makers choosing consistently, i.e. $E[y^*_i|c_{iz} = 1] = g(F^{-1}(E[c_{iz}]))$. In addition, infinite differentiability of $g$ and $F$ ensure imply the composite function $h \equiv g \circ F^{-1}$ will have a well-defined Taylor Series approximation of degree $D$. We then obtain 6.2 by integrating the marginal preference function $h(.)$ from $E[c_{iz}] = 0$ to $E[c_{iz'}]$ and scaling by $E[c_{iz'}]$ for any arbitrary $z'$. Finally, 6.3 follows directly from setting $E[c_{iz}] = 1$ in 6.2. Note that when $N = D$, we will have $D + 1$ equations in $D + 1$ unknowns, so that $a_0, ..., a_D$ are just-identified. When $N > D$, we will have more equations than unknowns, and a best-fit technique such as least squares can be used to estimate $a_0, ..., a_D$.

Figure 3.1 illustrates the approach with data from Table 3.3 under a linear functional form assumption: $E[y^*|c_i(z) = 1] = \alpha + \beta E[c_i(z)]$. With two values of $z$, $\alpha$ and $\beta$ are just-identified: $\alpha \approx 0.82$ and $\beta \approx -0.35$. Applying 6.2 yields $E[y^*_i] = 0.48$.

4. Application to 401(k) Automatic Enrollment

In this section we illustrate the identification framework developed above with data on enrollment into employer-provided 401(k) pension plans. Enrollment may be opt-in or opt-out, and an influential body of research documents striking differences in take-up and

---

98The approximation disregards terms of order $D$ and higher, i.e. those of the form $E[c_{iz}]^k$ where $k \geq D$, as do the approximations in (6.2) and (6.3).
Figure 3.1: Extrapolation from Preferences of Consistent and Marginally Consistent Choosers

- \( E[y^*] \approx 0.48 \)
- \( E[y^* | c_i = 1] \)
- \( E[y^* | z_i^* = z] \)

savings behavior between the two regimes (Madrian and Shea, 2001; Choi et al., 2006; Chetty et al., 2014). Although we are not the first to study the welfare implications of defaults in this setting (Carroll et al., 2009; Bernheim, Fradkin and Popov, 2014), an important advantage to our approach is that the preference information we recover does not require taking a stance on the exact positive model that generates the decision-making bias.

To apply our framework to this setting, we focus on the extensive margin of 401(k) participation, i.e., how automatic enrollment affects whether employees choose to participate. We use published aggregate data from Madrian and Shea (2001). These data come from a large healthcare and insurance firm which switched from an opt-in to an opt-out enrollment design in April 1998.\(^99\) Madrian and Shea (2001) compare enrollment rates for employees hired before and after the policy change; they document larger participation rates among employees hired after the switch to opt-out enrollment.

Let \( y_i \) indicate whether an employee enrolls in the plan and let \( d_0 \) and \( d_1 \) correspond to the frame being opt-in and opt-out, respectively. Frame separability seems likely to hold in this setting: it is difficult to imagine an employee’s preferences over how much to

---

\(^{99}\) Under automatic enrollment, employees who were automatically enrolled faced a default contribution rate of 3 percent.
save depend on how his employer chooses to structure enrollment into its sponsored retirement plan. In contrast, whether the consistency principle holds for all individuals is less certain. The consistency principle requires that employees who would make the same participation decision under both opt-in and opt-out enrollment must actually prefer the option that they choose. The assumption would be violated, for example, if some of those employees who would (consistently) choose not to participate in the plan under either policy were myopic, and would actually be better off if they were to participate. On the other hand, for purposes of policy design, it may be desirable to treat the consistent choices made in this setting as revealing employees’ preferences, based on a non-paternalistic principle of respecting choice Bernheim and Rangel (2009). Turning to unconfoundedness, the condition is satisfied so long as the identifying assumption in Madrian and Shea (2001) is correct, i.e., that an employee’s hiring date is uncorrelated with her enrollment preferences or her susceptibility to the frame. Finally, frame monotonicity requires that no employee chooses to enroll when enrollment is opt-in but chooses not to enroll when enrollment is opt-out, which seems plausible in this setting.

Table 3.4 reports the aggregated choice data by frame. We estimate \( Y_0 = 0.374 \) and \( Y_1 = 0.859 \). Substituting these values into the definition of \( Y_c \) from Proposition 1 yields \( Y_c = 0.726 \), with standard error 0.008. Thus under frame monotonicity, of the 51.5 percent of employees whose enrollment decisions were unaffected by the enrollment design, we may conclude that a large majority (72.6 percent) preferred enrollment in the plan. Without assuming frame monotonicity, Proposition 1.2 implies the fraction of unaffected employees preferring enrollment was at least 72.6 percent.

A tricky case occurs when default effects are driven by employees’ interpreting the default as advice from their employer. In that case, the frame affects the ex ante expected utility of the available options to the employee but not the ex post realized utility to the employee of enrollment or non-enrollment – the relevant parameters for designing the optimal policy. See Online Appendix C. Importantly, not all forms of present-bias cause the consistency principle to fail. For example, in the model of default-sensitivity studied by Carroll et al. (2009), present-bias causes individuals to procrastinate and stick with the default savings plan until they make an active choice, but when they do make an active choice, the amount they choose to save is optimal. Such behavior satisfies the consistency principle because those individuals who choose consistently (i.e., who select the same option under either default), have selected their welfare-maximizing option.
Turning to population preferences, the bounds provided by Proposition 2 are quite coarse given the large fraction of inconsistent decision-makers. With frame monotonicity, we may only conclude that population preferences lie somewhere between 0.374 and 0.859; without frame monotonicity, we may only rule out values of $E[y^*_i]$ below 0.233. Consequently, additional structure is needed before one may draw more precise welfare conclusions from the data.

To shed light on the correlation between employees’ preferences and their sensitivity to the enrollment regime, we utilize the matching-on-observables estimator described in Section 4.2. Madrian and Shea (2001) report disaggregated choice data by employee compensation, reproduced in the first two rows of Table 3.4. Computing consistent employee preferences and consistency rates by income yields a striking pattern, which is depicted in Panels A and B of Figure 3.2. The decisions of low-earning employees are much more likely to be affected by the default, and among those who are consistent, fewer prefer enrollment. We observe similar patterns when we conduct the analysis by age instead of income, as displayed in Panels C and D of Figure 3.2, and also when we disaggregate the data by race or gender.

These findings are easy to rationalize: it could be, for example, that younger, lower-income employees are inattentive to retirement savings because retirement is far off, and those who are attentive save less because they anticipate higher future earnings. In addition, these results point to a positive correlation between preferences and consistency, and hence a violation of decision quality independence. It suggests that those employees whose choices are sensitive to the default are less likely to prefer enrollment than those employees whose choices are unaffected.

Identifying the preferences of the inconsistent employees by matching on income re-

102 Although the present data do not allow us to implement them, one could also imagine a range of instrumental variables strategies that might be implemented in this setting, such as sending workers extra reminders about enrollment, reducing or increasing transaction costs associated with opting out of the default (e.g., lengthening or shortening the required forms); varying the enrollment sign up window, or providing some employees with financial counseling sessions before making their decisions.

103 The aggregate nature of the data prevents us from disentangling these relationships with a multivariate specification.
quires conditional decision quality independence – i.e., that among employees with the same income, sensitivity to the design of the enrollment decision is uncorrelated with preferences for enrollment. Although with richer microdata one would ideally account for other variables in addition to income – such as education or experience – which may also play a role, adjusting for income in this way is likely to be an improvement over estimates that impose unconditional decision quality independence.104

Table 3.4 presents the results of this analysis.105 Using Proposition 3, we estimate that 61.9 percent of inconsistent employees, and 66.9 percent of all employees, prefer enrollment. The difference in estimated preferences between the consistent and inconsistent employees is statistically significant, allowing us to reject the hypothesis of decision quality independence at the 1 percent confidence level ($p < 0.001$). Although a 60 percent majority of inconsistent employees prefer enrollment, the results suggest fewer than half of the lowest-income employees do so, implying that those in this group may be better off under an opt-in regime.

104 With microdata, one could estimate $E[y_{1i}|w]$ and $E[y_{0i}|w]$ with rich demographics $w$ using any of several techniques, such as multiple linear regression or kernel regression. Applying the identification results in Proposition 3 in this setting would mean using the estimated conditional expectations to calculate $Y_c(w)$ and $s_w$ for each value of $w$, and then taking weighted averages implied by 3.1 and 3.2 based on the empirical distribution of $w$, or for that matter any counterfactual distribution of $w$.

105 The published version of Madrian and Shea (2001) does not provide one final input for the calculation of population preferences using the matching estimator, which is the share of the population in each of the income bins. For the sake of illustration, we use hypothetical shares (reported in row 3 of Table 3.4), chosen to match aggregate enrollment rates and the mean and median compensation figures provided in the original study. Our results should therefore be interpreted as applying to a firm characterized by the income distribution we assume, which may turn out to differ from the actual firm studied by Madrian and Shea (2001).
Table 3.4: Application to Enrollment in 401(k) Plans, by Total Compensation

<table>
<thead>
<tr>
<th>Enrollment rate under opt-out</th>
<th>$&lt;20k</th>
<th>$20k-$29k</th>
<th>$30k-$39k</th>
<th>$40k-$49k</th>
<th>$50k-$59k</th>
<th>$60k-$69k</th>
<th>$70k-$79k</th>
<th>$&gt;80k</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>79.5</td>
<td>82.8</td>
<td>88.9</td>
<td>91.8</td>
<td>92.8</td>
<td>94.7</td>
<td>91.5</td>
<td>94.1</td>
<td>85.9</td>
</tr>
<tr>
<td>(1.19)</td>
<td>(0.84)</td>
<td>(1.06)</td>
<td>(1.61)</td>
<td>(1.52)</td>
<td>(1.32)</td>
<td>(1.64)</td>
<td>(0.97)</td>
<td>(0.46)</td>
<td></td>
</tr>
<tr>
<td>Enrollment rate under opt-in</td>
<td>12.5</td>
<td>24.5</td>
<td>42.2</td>
<td>51.0</td>
<td>61.6</td>
<td>59.7</td>
<td>57.9</td>
<td>68.3</td>
<td>37.4</td>
</tr>
<tr>
<td>(1.13)</td>
<td>(1.12)</td>
<td>(1.96)</td>
<td>(3.43)</td>
<td>(3.34)</td>
<td>(3.37)</td>
<td>(3.39)</td>
<td>(2.26)</td>
<td>(0.70)</td>
<td></td>
</tr>
<tr>
<td>Percent of employees</td>
<td>20.0</td>
<td>35.0</td>
<td>15.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>5.0</td>
<td>10.0</td>
<td>100</td>
</tr>
<tr>
<td>Percent consistent</td>
<td>33.0</td>
<td>41.7</td>
<td>53.3</td>
<td>59.2</td>
<td>68.8</td>
<td>65.0</td>
<td>66.4</td>
<td>74.1</td>
<td>51.5</td>
</tr>
<tr>
<td>(1.64)</td>
<td>(1.39)</td>
<td>(2.23)</td>
<td>(3.79)</td>
<td>(3.67)</td>
<td>(3.61)</td>
<td>(3.76)</td>
<td>(2.46)</td>
<td>(0.83)</td>
<td></td>
</tr>
<tr>
<td>Percent of consistent preferring enrollment</td>
<td>37.9</td>
<td>58.8</td>
<td>79.2</td>
<td>86.1</td>
<td>89.5</td>
<td>91.8</td>
<td>87.1</td>
<td>92.2</td>
<td>72.6</td>
</tr>
<tr>
<td>(2.53)</td>
<td>(1.61)</td>
<td>(1.76)</td>
<td>(2.48)</td>
<td>(2.04)</td>
<td>(1.91)</td>
<td>(2.25)</td>
<td>(1.23)</td>
<td>(0.80)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The first two rows of data come from Table IV of Madrian and Shea (2001) (N = 4249 without automatic enrollment, 5801 with). Population shares by compensation (row 3) are constructed to match aggregate enrollment rates and mean and median compensation figures provided in the original study. Calculations relying on these hypothetical data are marked by *. Standard errors are calculated assuming these hypothetical shares match the distribution of income within both enrollment regimes.

Figure 3.2: The Empirical Relationship between Consistency and Preference for Enrollment in a 401(k)

Panel A. Fraction of Consistent Preferring Enrollment, by Compensation

Panel B. Fraction Consistent, by Compensation

Panel C. Fraction of Consistent Preferring Enrollment, by Age

Panel D. Fractions Consistent, by Age

Source: Authors’ calculations using data from Madrian and Shea (2001). Panels A and B are identical to the fourth and seventh rows of Table IV. Panels C and D are calculated in the same way as Panels A and B using age instead of compensation.
5. Conclusion

Recovering preferences from choice data is a fundamental problem in behavioral economics; the presence of systematic “choice-reversals” casts doubt on the revealed preference approach that underlies neoclassical welfare analysis. We relax the standard revealed preference approach to accommodate the evidence that decision-makers sometimes choose differently based on preference-irrelevant features of the choice situation. Like Bernheim and Rangel (2009), there is a sense in which our relaxation of the standard approach is the minimum required to accommodate the observed inconsistencies, in that we assume that decision-makers who do choose consistently across frames are revealing their true preferences. By imposing additional structure on the problem in the form of a frame monotonicity assumption, the problem of preference recovery is transformed into a familiar problem of endogeneity: whether an individual reveals her preferences through choice may depend on her preferences over the objects being chosen. This transformed problem is not only more familiar, it is also more tractable: over the last 50 years, economists have developed a wide range of tools for dealing with endogeneity in the recovery of parameters of this sort. In this paper we have adapted a number of these tools to the problem of identifying preferences in the presence of inconsistent decision-making.

Although focusing on binary menus and binary frames simplifies the analysis, our framework is useful outside of such settings as well. Online Appendix D develops several generalizations to more complicated choice settings. Notably, a number of the present results extend in a straightforward way to ordered menus with two frames and multiple options. We also develop generalizations to settings with multi-dimensional frames or multiple frames that vary in their intensity.

An important feature of our approach is its reduced-form nature. Within the wide range of models consistent with frame-monotonicity and the consistency principle, the basic identification problem – i.e., understanding the empirical correlation between decision-makers’ preferences and their sensitivity to frames – is the same regardless of the details of the
structural model that generates behavior. On the other hand, our approach is not a replacement for structural models of decision-making. As in other areas of empirical economics, the interpretation of the parameters identified by reduced-form approaches depend on the underlying structural model that generates behavior. As described in Online Appendix C, understanding the underlying structural model provides guidance about which types of control variables are needed for conditional decision quality independence to hold and about which types of variation constitute valid decision quality instruments.

The framework studied here can be thought of as a special case of a more general approach, in which an observer first identifies the preferences of a reference group of decision-makers whose choices are assumed to reveal their true preferences, and subsequently extrapolates those preferences to the rest of the population. In our approach, the reference group consists of those decision-makers who choose consistently across frames. The benefits of our approach are that it allows us to avoid ex ante assumptions about which particular decision-makers optimize, and it exhibits deference to decision-makers’ (coherent) choice behavior. On the other hand, choosing the reference group in other ways will be desirable in settings where there is reason to be skeptical that the decisions of the consistent decision-makers actually reveal their true preferences. Instead of consistent decision-makers, the reference group might consist of experts, experienced choosers, or those thought to be immune to the framing effect in question (e.g., Johnson and Rehavi, 2013; Bronnenberg et al., 2013; Handel and Kolstad, 2015). The identification techniques we have proposed may be utilized with these reference groups as well; for example, one might want to adjust the recovered preferences of experts based on observable characteristics before extrapolating their preferences to the rest of the population, or utilize instrumental variation that causes some individuals to become experts.

The methods described here are subject to important limitations. First, in certain applications, consistent choices may not in fact reveal preferences. For example, even decision-makers who consistently choose one retirement plan over another, regardless of the default
option, may still be choosing sub-optimally based on, for example, present bias. Similarly, biases in judgment and perception – such as over-optimism or a tendency to underweight low-risk events – may manifest themselves consistently across frames. Many of these failures can be attributed within our framework to the presence of “missing” frames, which affect behavior but do not vary in the data available to the researcher. Accurately identifying preferences in such contexts requires additional data or assumptions that permit the analysis to move further away from observed choice behavior.

Second, although we have attempted to develop identification strategies that may be applied to data, the credibility of such strategies will turn on whether their assumptions are met in the application at hand. Insofar as one is skeptical that the required assumptions will be satisfied in any setting, our results highlight the difficulty in conducting even weakened forms of revealed reference analysis in the presence of framing effects. Further work, perhaps combining the current framework with data on subjective well-being, could attempt to empirically assess the validity of the underlying assumptions about welfare made here.

A third limitation is that the ordinal preferences over menu objects that our approach identifies may not be the only preferences that are welfare-relevant in a particular application. For example, the analysis in Online Appendix B noted conditions in which the optimal choice of frame will depend on the relative magnitude of utility costs incurred by choosing against the frame. As in non-behavioral settings, identifying cardinal preference information from choices of the type we presume to observe would require additional data or richer structure than what we impose here. Similarly, in some contexts policy decisions about frames will shape decisions that generate externalities, such as opt-in versus opt-out rules for organ donations (Abadie and Gay, 2006; Johnson and Goldstein, 2003) or, in the environmental context, plastic bag taxes versus subsidies (Homonoff, 2014). In those applications, the private preferences of decision-makers, while still important, will not be the only relevant parameters for setting the optimal policy. Additionally, determining the optimal frame will often require non-ordinal information regarding the intensity of incon-
sistent decision-makers’ preferences, a task we briefly discuss in Online Appendix C.2. To summarize, we have provided methods for identifying one type of normatively-relevant preference information; in some contexts, other types of preference information will be relevant as well. Developing methods to recover that information and incorporate it into optimal policy prescriptions is an important task for future research.

Appendix

A. Proofs

Proof of Proposition 1

The proof of 1.1 is provided in the body of the paper.

To prove 1.2, let \( \alpha = p(y_{i0} = 1; y_{i1} = 0) \) denote the fraction of frame defiers and note that now \( p(c_i = 0) = p(y_{i1} = 1; y_{i0} = 0) + \alpha \). It is straightforward to show that

\[
E[y_{i0}] = p(c_i = 1)E[y^*_i | c_i = 1] + \alpha \quad (3.9)
\]

\[
E[y_{i1}] = p(c_i = 1)E[y^*_i | c_i = 1] + p(c_i = 0) - \alpha \quad (3.10)
\]

Substituting these into the definition of \( Y_c \), we have

\[
Y_c = \frac{p(c_i = 1)E[y^*_i | c_i = 1] + \alpha}{p(c_i = 1) + 2\alpha} \quad (3.11)
\]

Subtract \( \frac{1}{2} \) from both sides of (3.11) to obtain

\[
Y_c - \frac{1}{2} = \frac{p(c_i = 1)(E[y^*_i | c_i = 1] - \frac{1}{2})}{p(c_i = 1) + 2\alpha} \quad (3.12)
\]

In addition, subtracting \( E[y^*_i | c_i = 1] \) from both sides of (3.11) yields

\[
Y_c - E[y^*_i | c_i = 1] = \frac{\alpha(1 - 2E[y^*_i | c_i = 1])}{2\alpha + p(c_i = 1)} \quad (3.13)
\]
This expression gives the bias in $Y_c$ when frame monotonicity fails. To complete the proof, note that (3.12) implies that $Y_c \geq \frac{1}{2} \iff E[y^*|c_i = 1] \geq \frac{1}{2}$, and (3.13) implies that $E[y^*|c_i = 1] \geq \frac{1}{2} \iff E[y_i^*|c_i = 1] \geq Y_c$.

Proof of Proposition 2

By the law of iterated expectations, we can write:

$$E[y_i^*] = E[y_i^*|c_i = 1]p(c_i = 1) + E[y_i^*|c_i = 0]p(c_i = 0)$$

(3.14)

First we assume frame monotonicity to prove 2.1. In the proof of proposition 1.1 we showed that $E[y_i^*|c_i = 1]p(c_i = 1) = p(y_i^* = 1; c_i = 1) = Y_0$, and $p(c_i = 0) = Y_1 - Y_0 > 0$. Substituting these into (3.14) yields

$$E[y_i^*] = Y_0 + E[y_i^*|c_i = 0](Y_1 - Y_0)$$

(3.15)

Proposition 2.1 follows from the fact that (3.15) is strictly monotonic in $E[y_i^*|c_i = 0]$ and $E[y_i^*|c_i = 0] \in [0, 1]$.

To prove 2.2, note that by (3.9) and (3.10),

$$p(c_i = 1) = Y_0 + 1 - Y_1 - 2\alpha$$

(3.16)

$$E[y_i^*|c_i = 1] = \frac{Y_0 - \alpha}{Y_0 + 1 - Y_1 - 2\alpha}$$

(3.17)

Note that (3.17) implies that consistent preferences are point-identified when the prevalence of frame defiers, $\alpha$, is known. Substituting (3.16) and (3.17) into (3.14) yields

$$E[y_i^*] = Y_0 - \alpha + E[y_i^*|c_i = 0](Y_1 - Y_0 + 2\alpha)$$

(3.18)

Because $p(c_i = 0) = Y_1 - Y_0 + 2\alpha \geq 0$, this expression is increasing in $E[y_i^*|c_i = 0]$.

For the lower bound, set $E[y_i^*|c_i = 0] = 0$: $E[y_i^*] \geq Y_0 - \alpha$. This expression is decreasing in $\alpha$, so we obtain a lower bound with the maximum possible $\alpha$. By (3.17), $E[y_i^*|c_i = 1] \leq 1$.
implies $\alpha \leq 1 - Y_1$. It follows that $E[y^*_i] \geq Y_0 - (1 - Y_1)$, which is only binding when
$Y_0 - (1 - Y_1) \geq 0$.

For the upper bound, set $E[y^*_i|c_i = 0] = 1$ in (3.18): $E[y^*_i] \geq Y_1 + \alpha$. This expression
is increasing in $\alpha$, so we obtain an upper bound by setting the maximum possible $\alpha$. By
(3.17), $E[y^*_i|c_i = 1] \geq 0$ implies $\alpha \leq Y_0$.\footnote{Combining insights from these two cases, it follows that $\alpha \leq \min\{Y_0, 1 - Y_1\}$. Which of these two
constraints is binding determines whether we obtain an upper or a lower bound for $E[y^*_i]$.} It follows that
$E[y^*_i] \leq Y_0 + Y_1$, and this upper bound is binding whenever $Y_0 + Y_1 \leq 1$.

Proof of Lemma 1

Throughout the proofs of Lemma 1 and Proposition 3 we suppress the notation for conditioning on $w_i = w$, so that, for example, $E[y^*_i|w] \equiv E[y^*_i|w_i = w]$.

To prove L1.1, note that Bayes Rule implies

$$p(\{\{w_i = w|c_i = 1\} = \frac{p(c_i = 1|w)}{p(c_i = 1)}p(w_i = w)$$

In the proof of Proposition 1.1, we showed that $p(c_i = 1) = Y_0 + 1 - Y_1$ under uncondi-
tional unconfoundedness (A3). Repeating the proof of Proposition 1 while conditioning on $w$, under conditional unconfoundedness (A3’), yields $p(c_i = 1|w) = Y_0(w) + 1 - Y_1(w)$. By the law of total probability, $p(c_i = 1) = E_w[Y_0(w) + 1 - Y_1(w)]$. Substituting these two
expressions into (3.19) yields $p(\{\{w_i = w|c_i = 1\} = q_w p_w$.

The proof that $p(\{\{w_i = w|c_i = 0\} = s_w p_w$ is analogous.

Proof of Proposition 3

To prove 3.1, note that by the law of iterated expectations

$$E[y^*_i] = E_w[E[y^*_i|w]]$$

106Combining insights from these two cases, it follows that $\alpha \leq \min\{Y_0, 1 - Y_1\}$. Which of these two
constraints is binding determines whether we obtain an upper or a lower bound for $E[y^*_i]$.
Repeating the proof of Proposition 1.1 while conditioning on \( w \) yields \( E[y^*_{i} | c_i = 1, w] = Y_c(w) \). Conditional decision quality independence (A5) implies \( E[y^*_{i} | w] = Y_c(w) \). Substituting this into (3.20) yields the desired result.

To prove 3.2, we begin by observing that

\[
E[y^*_{i} | c_i = 0] = \frac{p(y^*_{i} = 1; c_i = 0)}{p(c_i = 0)}
\]

Applying the law of iterated expectations to the numerator yields

\[
E[y^*_{i} | c_i = 0] = E_w[p(y^*_{i} = 1; c_i = 0 | w)] \tag{3.21}
\]

By the definition of conditional probability, we know that for any \( w \)

\[
p(y^*_{i} = 1; c_i = 0 | w) = E[y^*_{i} | c_i = 0, w] p(c_i = 0 | w)
\]

As above, conditional decision quality independence (A5) implies \( E[y^*_{i} | c_i = 0, w] = E[y^*_{i} | w] = Y_c(w) \). Equation (3.21) then implies\(^{107} \) that

\[
y^*_{i} | c_i = 0 = E_w \left[ \frac{p(c_i = 0 | w)}{p(c_i = 0)} Y_c(w) \right]
\]

Substituting the result from L1.2 that \( \frac{p(c_i = 0 | w)}{p(c_i = 0)} = p(w_i = w | c_i = 0) = s_w \) yields the desired result.

\[\blacksquare\]

**Proof of Proposition 4**

By decision quality monotonicity (A7), we can divide the population into three groups based on \( (c_{ih}, c_{il}) \): the always consistent (A) with \( c_{ih} = c_{il} = 1 \), the sometimes consistent (S) with \( c_{ih} = 1; c_{il} = 0 \), and the never consistent (N) with \( c_{ih} = c_{il} = 0 \). Let \( \pi_j \) denote the share of the population in each group for \( j = A, S, N \), and let \( E[y^*_{i} | j] \) denote the fraction of each group preferring \( y^* \).

\(^{107}\)The fact that \( p(c_i = 0) \) is constant with respect to \( w \) allows us to move it inside the expectations operator in this expression.
For each fixed $z = z_k$, the conditions are identical to those in Proposition 1.1. Following the same logic as in the proof of Proposition 1.1, we have $Y_{0k} = p(y^* = 1; c_{ik} = 1) = E[y^* | c_{ik} = 1] p(c_{ik} = 1)$.

At each given $z$ we know which groups are consistent, so we know that

$$Y_0l = E[y^*_i | A] \pi_A$$

$$Y_{0h} = E[y^*_i | A] \pi_A + E[y^*_i | S] \pi_S$$

$$Y_{0h} - Y_{0l} = E[y^*_i | S] \pi_S$$

(3.22)

We also showed in the proof of Proposition 1.1 that $p(c_{ik} = 1) = Y_{0l} + 1 - Y_{0h}$. It follows that

$$p(c_{il} = 1) = \pi_A = Y_{0l} + 1 - Y_{1l}$$

$$p(c_{ih} = 1) = \pi_A + \pi_S = Y_{0h} + 1 - Y_{1h}$$

$$\pi_S = Y_{1l} - Y_{0l} - (Y_{1h} - Y_{0h})$$

(3.23)

Dividing (3.22) by (3.23) yields the desired result.

**Proof of Proposition 5**

Throughout the proof we let $\bar{c}(z) = E[c_{iz}]$. Fix any $D \in Z^+$. Our technical assumptions – requiring $F(z) = 0$, $F(\bar{z}) = 1$, and $F$ strictly increasing – imply that $F$ has a well-defined inverse function over the unit interval $[0, 1]$. Because we have assumed $F(z)$ and $g(z)$ are continuous and infinitely differentiable, the function $h = g \circ F^{-1}$ will be continuous and infinitely differentiable as well. As a result, it has a well-defined Taylor Series approximation of degree $D$ about any point in $(0, 1)$. Noting that $h(\bar{c}(z)) = E[y^*_i | z^*_i = F(\bar{c}(z))] = E[y^*_i | z^*_i = z]$ proves 5.1.

Now note that the preferences of the consistent choosers at some $z'$, $E[y^*_i | c_{iz'} = 1] = E[y^*_i | z^*_i \leq z']$, can be expressed using the definition of conditional probability as

$$E[y^*_i | c_{iz'} = 1] = \frac{\int_{z=\bar{z}}^{z'=z'} g(z) f(z) dz}{F(z')}$$
We employ a change of variables, letting $\bar{c} = F(z)$, $d\bar{c} = f(z)dz$. From above, $g(z) = h(\bar{c}(z))$, so we obtain

$$E[y^*_i | c_{iz} = 1] = \int_{\bar{c}=0}^{\bar{c}=F(z')} \frac{h(\bar{c}) d\bar{c}}{F(z')}$$

Substituting our approximation for $h(\bar{c})$ into 5.1, evaluating the integral in the numerator, and dividing by $F(z') = \bar{c}(z')$ yields the desired result in 5.2.

The result in 5.3 follows from evaluating the expression in 5.2 at $\bar{c} = 1$. ■

### B. Optimal Choice of Frame

This section motivates the parameters we focus on in the body of the paper by highlighting their relevance for setting the optimal policy. In this Section we derive formulas for the optimal frame as well as for the optimal decision-quality state (when the decision-quality state is a choice variable for the policymaker). The model we consider is simple but appealing in that the welfare conclusions are robust to a range of alternative positive models (in the spirit of Chetty (2009b)).

This section derives three results. First, when decision-makers’ welfare depends solely on the option they end up selecting, the optimal frame depends solely on the preferences of the inconsistent decision-makers. Intuitively, the choice of frame does not affect the choices made by the consistent decision-makers and consequently the planner should ignore the preferences of that group when determining the optimal frame.

Second, when decision-makers experience normatively-relevant opt-out costs from choosing against the frame – for example by selecting an option other than the default – the preferences of the consistent decision-makers become relevant as well. In particular, when the planner’s goal is to maximize the fraction of decision-makers that select their most-preferred option (e.g., when preference intensity is homogenous) and minimize the fraction incurring an opt-out cost, the optimal frame depends on the weighted average of consistent and inconsistent decision-makers’ ordinal preferences, where the weights depend on the
magnitude of the opt-out costs and the fraction of consistent choosers in the population.

Third, we consider the problem faced by a planner who must decide whether to adopt a (potentially more expensive) decision-quality state, i.e., one in which more decision-makers will choose consistently. We show that the social welfare benefits achieved by improving individuals’ decision-making in this way depend on the difference in preferences between the decision-makers who are inconsistent at the high decision-quality state and the decision-makers who would be induced to choose consistently by the policy change. Intuitively, when this difference is large, a greater fraction of the sometimes-consistent decision-makers benefit from the increase in the decision-quality state and the social planner may provide the never-consistent with a better-tailored frame.

Setup

Assume a continuum population of measure 1 chooses from a fixed menu \( X = \{0, 1\} \). A benevolent planner selects which option is favored by the frame, \( d \in \{0, 1\} \). The decision quality state is given by \( z \in \mathbb{Z} \), which could be fixed (if \( \mathbb{Z} \) is a singleton set) or set by policy, with some associated implementation cost \( \kappa(z) \).\(^{108}\) We assume that the planner’s objective is to maximize the probability that individuals choose their preferred option, and possibly also to minimize the probability that they choose the option not favored by the frame (due to the presence of opt-out costs, for example). For simplicity, the relative weight the planner attaches to these two objectives is described by a single parameter, \( \gamma \). Formally, the planner’s problem is

\[
\max_{d \in \{0, 1\}, z \in \mathbb{Z}} \int_i \left[ 1\{y_{id} = y_i^d\} - \gamma 1\{y_{id} \neq d\} \right] di - \kappa(z)
\]

where \( 1\{\} \) is an indicator function equal to 1 when the expression inside the brackets is true and zero otherwise. This simple objective function corresponds to the case in which

\(^{108}\)The units of \( \kappa(z) \) is the number of individuals the planner would need to give their preferred option to justify incurring a cost of \( \kappa(z) \).
the planner seeks to maximize social welfare with homogeneous preference intensity, opt out costs, and pareto weights across individuals. In the more general setting, heterogeneity in these variables is reflected in the solution to the optimal policy problem. These restrictions require that an agent ending up with her preferred option contributes equally to social welfare whether her preferred option is 0 or 1, and that opting out is equally costly for all decision-makers who opt out, regardless of their preference or the default. \footnote{With a different social welfare function, such as one that incorporates the intensity of each individual’s preference for the available options, the parameters we identify will be insufficient to fully characterize optimal policy without additional restrictions. In general, identifying the intensity of individuals’ preferences requires richer data and/or stronger positive assumptions than the type we employ here. In particular, implementing the solution to a utilitarian planner’s problem would require estimating the joint distribution of (1) decision-makers’ cardinal preferences and (2) the welfare cost to a decision-maker of choosing against the frame.}

**Results**

We prove two simple propositions characterizing the solution to the planner’s problem. The first considers the optimal frame when the decision-quality state is fixed. The second considers the joint choice of the optimal frame and the optimal decision-quality state, assuming for simplicity that there are no opt-out costs.

**Proposition A.1**

Suppose $Z$ is singleton, $Z = \{z\}$. Assume the consistency principle (A2) and frame monotonicity (A4).\footnote{The assumption of frame separability is embedded in the planner’s problem. Unconfoundedness is necessary for the statistical estimation of the relevant preference parameters from data, but not for understanding the relationship between these parameters and optimal policy, our focus here.} Let $\bar{c} = E[c_{iz}]$. Let $\rho = \frac{\bar{c}}{\bar{c} + 1 - \bar{c}}$. The optimal frame is $d_1$ if and only if

$$
(1 - \rho)E[y^*_i | c_i = 0] + \rho E[y^*_i | c_i = 1] > \frac{1}{2}
$$

\hspace{1cm} (3.24)

**Proof:** Note that the planner’s problem above is equivalent to the following

$$
\max_{d \in \{1,0\}, z \in Z} p(y_{id} = y^*_i) - \gamma p(y_{id} \neq d) - \kappa(z)
$$
Since $z$ is fixed by assumption, the solution to the planner’s problem simplifies to the comparison of the objective function evaluated at $d = 1$ and $d = 0$. We will therefore have that $d = 1$ is superior if and only if

$$\frac{p(y_{i1} = y^*) - \gamma p(y_{i1} = 0)}{p(y_{i0} = y^*) - \gamma p(y_{i0} = 1)} > \frac{1}{2}$$  \hfill (3.25)

We next derive each of these probabilities.

By frame monotonicity, the consistency principle, and the law of iterated expectations, terms one and three simplify to:

$$p(y_{i1} = y^*_i) = p(c_i = 1) + E[y^*_i | c_i = 0]p(c_i = 0)$$

$$p(y_{i0} = y^*) = p(c_i = 1) + E[1 - y^*_i | c_i = 0]p(c_i = 0)$$

By frame monotonicity, terms two and four simplify to:

$$\gamma p(y_{i1} = 0) = \gamma p(y_{i1} = y_{i0} = 0) = \gamma E[1 - y^*_i | c_i = 1]p(c_i = 1)$$

$$\gamma p(y_{i0} = 1) = \gamma p(y_{i1} = y_{i0} = 1) = \gamma E[y^* | c_i = 1]p(c_i = 1)$$

Combining terms and simplifying, we have that $d = 1$ is optimal if and only if

$$E[y^*_i | c_i = 0] \frac{p(c_i = 0)}{\gamma p(c_i = 1) + p(c_i = 0)} + E[y^*_i | c_i = 1] \frac{\gamma p(c_i = 1)}{\gamma p(c_i = 1) + p(c_i = 0)} > \frac{1}{2}$$

Substituting the definitions of $\rho$ and $\gamma$ yields the desired result.

**Discussion of Proposition A.1** The optimal frame will tend toward $d = 1$ when (1) the fraction of inconsistent choosers who prefer $y = 1$ is large, and (2) the fraction of consistent choosers who prefer $y = 1$ is large. The first group is helped by the frame being $d = 1$, because the default directly influences their choices. The second group is helped by the frame being $d = 1$, because they will not incur an opt-out cost to receive their preferred option. The left-hand side of (3.24) shows that the relative importance of the consistent

---

130
group is determined by a weight $\rho$, which is increasing in opt-out costs $\gamma$ and the size of the consistent subgroup ($c$). Note that when $\gamma = 0$, $\rho = 0$, so the condition for optimality of $d = 1$ simplifies to $E[y_i^*|c_i = 0] > \frac{1}{2}$. Finally, note that the optimal policy does not depend on $\gamma$ when the majority of consistent and inconsistent decision-makers prefer the same option.

**Proposition A.2**

Suppose the planner can choose between two decision quality states, $Z = \{z_h, z_l\}$. Suppose $\kappa(z^h) > \kappa(z^l)$, and let $\Delta \kappa = \kappa(z^h) - \kappa(z^l)$ be the change in per-person cost of increasing the decision-quality state from $z_l$ to $z_h$. Assume the consistency principle (A2), frame monotonicity (A4), decision-quality exclusivity (A6), decision quality monotonicity (A7), and that $\gamma = 0$. Then the solution to the planner’s problem is given case-wise by

1. $(1, z_l)$ if
   
   (a) $Y^*_N > \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} > \frac{1}{2}$, and $\Delta \kappa > (1 - Y^*_S) \pi_S$, OR
   
   (b) $Y^*_N > \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} < \frac{1}{2}$ and $\Delta \kappa > Y^*_S \pi_S + (2Y^*_N - 1) \pi_N$

2. $(0, z_l)$ if
   
   (a) $Y^*_N < \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} < \frac{1}{2}$ and $\Delta \kappa > Y^*_S \pi_S$, OR
   
   (b) $Y^*_N < \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} > \frac{1}{2}$ and $\Delta \kappa > \pi_S (1 - Y^*_N) + (1 - 2Y^*_N) \pi_N$

3. $(0, z_h)$ if
   
   (a) $Y^*_N < \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} < \frac{1}{2}$ and $\Delta \kappa < Y^*_S \pi_S$, OR
   
   (b) $Y^*_N > \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} < \frac{1}{2}$ and $\Delta \kappa < Y^*_S \pi_S + (2Y^*_N - 1) \pi_N$

4. $(1, z^h)$ if
   
   (a) $Y^*_N > \frac{1}{2}$ and $\frac{Y^*_N \pi_N + Y^*_S \pi_S}{\pi_N + \pi_S} > \frac{1}{2}$ and $\Delta \kappa < (1 - Y^*_S) \pi_S$, OR
(b) \( Y_N^* < \frac{1}{2} \) and \( \frac{Y_N^* \pi_N + Y_S^* \pi_S}{\pi_N + \pi_S} > \frac{1}{2} \) and \( \Delta \kappa < (1 - Y_N^*) \pi_S + (1 - 2Y_S^*) \pi_N \)

where \( Y_N^* \equiv E[y_i^*|c_{ih} = c_{il} = 0] \), \( Y_S^* \equiv E[\phi_i|c_{ih} > c_{il}] \), \( \pi_N = 1 - E[c_{ih}] \), and \( \pi_S = E[c_{ih}] - E[c_{il}] \).

**Proof:** When \( \gamma = 0 \) the planner’s objective evaluated at each of the four possible \( d \) by \( z_k \) combinations is

\[
p(y_{idk} = y_i^* - \kappa(z_k))
\]

In the proof of Proposition A.1, we showed that these four expressions can be re-written for fixed \( z_k \) as

\[
d = 1: \quad p(c_{ik} = 1) + E[y_i^*|c_{ik} = 0]p(c_{ik} = 0) - \kappa(z_k)
\]

\[
d = 0: \quad p(c_{ik} = 1) + E[1 - y_i^*|c_{ik} = 0]p(c_{ik} = 0) - \kappa(z_k)
\]

By decision-quality monotonicity and the existence of consistent choosers, we can divide the population into always-consistent (A), never-consistent (N) and sometimes-consistent (C), exactly as in Propositions 4 and A.1. The fraction preferring \( y = 1 \) in each population are given by \( Y_A^* \), \( Y_N^* \), and \( Y_S^* \), respectively, and the size of each population is given by \( \pi_A \), \( \pi_N \), and \( \pi_S \), respectively.

Using this notation, \( p(c_{il} = 1) = \pi_A \), \( p(c_{ih} = 0) = \pi_N \), \( p(c_{il} = 0) = 1 - \pi_A = \pi_N + \pi_S \). By the law of iterated expectations and the definition of conditional probability,

\[
E[y_i^*|c_{il} = 0] = \frac{Y_S^* \pi_S + Y_N^* \pi_N}{\pi_S + \pi_N}
\]

\[
E[1 - y_i^*|c_{il} = 0] = \frac{(1 - Y_S^*) \pi_S + (1 - Y_N^*) \pi_N}{\pi_S + \pi_N}
\]

Using these expressions to simplify equations (3.26) and (3.27) for both values of \( z_k \) yields:

\[
d = 1, z = z_l: \quad \pi_A + Y_S^* \pi_S + Y_N^* \pi_N - \kappa(z_l)
\]

\[
d = 0, z = z_l: \quad \pi_A + (1 - Y_S^*) \pi_S + (1 - Y_N^*) \pi_N - \kappa(z_l)
\]

132
\[
\begin{align*}
d = 1, z = z_h : & \quad \pi_A + \pi_S + Y_N^*\pi_N - \kappa(z^h) \\
d = 0, z = z_h : & \quad \pi_A + \pi_S + (1 - Y_N^*)\pi_N - \kappa(z^h)
\end{align*}
\] (3.30) (3.31)

Note that the first three terms in each of these will be the total number of individuals who receive their preferred option when the planner chooses that \((d, z)\) combination.

First, consider situations where the planner chooses \(d = 1\) regardless of \(z\). This requires \((3.28) > (3.29)\) and \((3.30) > (3.31)\), which simplify to the first two conditions in (1a) and (4a). The planner will set \(z_h\) if \((3.30) > (3.28)\), which simplifies to \(\Delta \kappa < (1 - Y_S^*)\pi_S\), which yields (4a). With the inequality reversed, we get (1a).

Second, consider situations where the planner chooses \(d = 0\) regardless of \(z\). This requires \((3.28) < (3.29)\) and \((3.30) < (3.31)\), which simplify to the first two conditions in (2a) and (3a). Then the planner chooses \(z_h\) if \((3.31) > (3.29)\), which simplifies to \(\Delta \kappa < Y_S^*\pi_S\). This yields the final condition in (2a) and (3a).

Third, consider the situation where the planner would want to choose \(d = 1\) under \(z_h\) and \(d = 0\) under \(z_l\). This requires \((3.28) < (3.29)\) and \((3.30) > (3.31)\), which provides the first two conditions in (1b) and (3b). In this situation, the planner chooses \(z^h\) if \((3.30) > (3.29)\) and \(z^l\) otherwise. Performing this comparison, we have that the planner chooses \(z^h\) if \(\Delta \kappa < Y_S^*\pi_S + (2Y_N^* - 1)\pi_N\), which is the final condition in (3b). When the inequality is reversed, we obtain the final condition in (1b).

Finally, consider the situation where the planner would want to choose \(d = 0\) under \(z_h\) and \(d = 1\) under \(z_l\). This requires \((3.28) > (3.29)\) and \((3.30) < (3.31)\), which provide the first two conditions in (2b) and (4b). In this situation, the planner chooses \(z_h\) if \((3.31) > (3.28)\). Comparing these, we see that the planner chooses \(z^h\) if \(\Delta \kappa < (1 - Y_S^*)\pi_S + (1 - 2Y_N^*)\pi_N\), which is the final condition in (4b). When the inequality is reversed, we obtain the final condition in (2b).

The planner should switch to \(z_h\) from \(z_l\) if the number of individuals who receive their preferred option increases by enough to justify the increase in implementation cost \(\Delta \kappa\). We divide the cases first based on two possibilities for the solution to the problem: the optimal
choice of frame either depends on the choice of decision-quality environment or it does not. When the optimal choice of frame does not depend on \( z \), switching to \( z_h \) from \( z_l \) helps only those individuals who are consistent at \( z_h \) but not \( z_l \) (group S), and who prefer the option not associated with the frame. In other words, when the never-consistent choosers prefer the same frame as the sometimes- and never-consistent choosers together, the planner should only take into account how many sometimes-consistent choosers will receive their preferred option under the improved decision-quality state. Parts (1a), (2a), (3a), and (4a) correspond to these possibilities. In the second set of possibilities, the optimal frame changes as the planner increases from \( z_l \) to \( z_h \). This case obtains if the individuals who become consistent at \( z_h \) (group S) have different preferences from the group who never choose consistently (group N). Here, moving from \( z_h \) to \( z_l \) not only gives individuals in group S who prefer the non-framed option their preferred option, but it also allows the planner to set a better default for group N. For example, suppose the planner would want to set \( d = 0 \) in \( z_l \) but \( d = 1 \) in \( z_h \). This corresponds to Parts (1b) and (3b) of the proposition. Then we must have that \( Y_{S}^{\ast} < \frac{1}{2} \), \( Y_{N}^{\ast} > \frac{1}{2} \), and the preferences of the S group dominate when determining optimal policy under \( z_l \), which occurs if there are more of them or their preferences are more homogenous. In this case the benefit of switching to \( z_h \) includes not only the benefit of giving those in group S who prefer the non-framed option their preferred option, but also the benefit of setting a frame in accordance with the preferences of the remaining group who are never consistent, group N. How large this benefit is depends on how far \( Y_{N}^{\ast} \) is from \( \frac{1}{2} \) (i.e. how bad the \( z_l \) frame was for this group) as well as the size of group N.

C. Alternative Positive Models of Default Sensitivity

In this Appendix we consider the relationship between our framework and positive models of framing effects. In many cases, a range of alternative models of framing effects will be observationally equivalent given the available data, meaning that any one of them might
explain decision-makers’ observed sensitivity to the frame in question. As emphasized by Bernheim (2009), such model uncertainty poses a challenge for welfare analysis, since the preferences implied by a decision-maker’s observed choice behavior depend on the positive model of behavior that maps the decision-maker’s preferences into his or her choices. An important advantage of our approach is that it can shed light on decision-makers’ preferences without specifying the exact positive model of behavior that generates the observed framing effect, at least within the class of positive models consistent with the assumptions set out in the body of the paper (such as frame monotonicity and the consistency principle).

To illustrate, we consider the application of our framework to alternative behavioral models that have been proposed to explain why decision-makers exhibit sensitivity to default effects. An important lesson of this exercise is that our reduced-form framework is not a replacement for structural modeling. That is, although the framework is sufficiently general to accommodate many of the models we consider, the specific positive model shapes the interpretation of the reduced-form parameters that our proposed approaches recover and has important implications for which of the proposed approaches are likely to succeed in a given setting.

In each model, we assume that every decision-maker (DM) $i$ chooses from a fixed menu $X = \{0, 1\}$. DM’s valuations of the two options are given by $u_i(0)$ and $u_i(1)$, with the difference denoted by $\bar{u}_i = u_i(1) - u_i(0)$. We assume that the distribution of $\bar{u}_i$ is described by cumulative density function $F(\cdot)$.\textsuperscript{111} We denote the default by $d \in \{d_0, d_1\}$, where the subscript indicates which option is the default. We continue to denote $y^*_i = 1\{u_i(y) > u_i(x)\}$ and $c_i = 1\{y_{i1} = y_{i0}\}$.

\textsuperscript{111}Our focus in the body of the paper is only with decision-makers’ ordinal preferences, but in some of the models considered here, differences in preference intensity will explain some of the variation in decision-makers’ consistency.
C.1 Inattention

One reason why people might be drawn to the default option is that they do not pay attention to the other menu items. Here, we develop a model of default sensitivity based on inattention, using the approach of Masatlioglu, Nakajima and Ozbay (2012).

C.1.1 Setup

In the model of Masatlioglu, Nakajima and Ozbay (2012), one assumes that an individual pays attention only to some subset of the menu \( X \), but that she maximizes her preferences over the alternatives that she considers.\(^{112} \) In order to incorporate framing effects, we must specify an attention filter \( \Gamma \) which depends on \( d \). The attention filter \( \Gamma \) is a mapping from \((X, d)\) to a subset of \( X \), \( \Gamma(X, d) \subseteq X \).\(^{113} \) Given a utility function representation of individual \( i \)'s preferences, \( u_i(.) \), we can write the consumer’s choice as the solution to the utility-maximization problem restricted to \( \Gamma_i(X, d_i) \).

\[
y_i(d_i) = \arg \max_{y \in \Gamma_i(X, d_i)} u_i(y)
\]

(3.32)

Claim: When \( X \) is binary, frame monotonicity and the consistency principle will be satisfied if the individual always pays attention to the default option. Formally, the sufficient condition is

\[
\forall i, \forall j \in \{0, 1\} \; j \in \Gamma_i(X, d_j)
\]

(3.33)

Proof: Suppose condition (3.33) is satisfied. Recall that frame monotonicity is only violated when \((y_{i0}, y_{i1}) = (1, 0)\). Suppose that \( y_{i0} = 1 \). Then \( 1 \in \Gamma_i\{x, d_0\} \) by (3.32). By (3.33),

---

\(^{112}\)Instead of reflecting inattention to a subset of the menu, an alternative possibility for why some decision-makers tend to select the default option is that they follow the default as a decision-making heuristic, or shortcut. The model is formally equivalent to the limited attention model for binary menus, but the two models depart with more than two options. That is, the heuristic model is a special case of the limited attention model where the attention filter consists solely of a single option (whichever option happens to be the default).

\(^{113}\)Masatlioglu et al.’s key assumption is that \( \forall X, \; \Gamma_i(X, d) = \Gamma_i(X \setminus x, d) \) whenever \( x \notin \Gamma_i\{X, d\} \). This assumption is not directly relevant to our setting because our focus is on binary choices. However, in the non-binary case it will place additional restrictions on when preferences are revealed by choices.
By (3.32) again then, \( u_i(1) > u_i(0) \). Finally, \( 1 \in \Gamma_i\{x,d_1\} \) by (3.33), so we know that \( y_{i1} = 1 \). This \( y_{i0} = 1 \) \( \implies \) \( y_{i1} = 1 \), \( c_i = 1 \), and \( y_i^* = 1 \). Using a similar set of steps, we know that \( y_{i1} = 0 \) \( \implies \) \( y_{i0} = 1 \), \( c_i = 1 \), \( y_i^* = 0 \).

Intuitively, when DMs choose \( y = 1 \) under \( d = 0 \), they are “revealing” that they pay attention to \( y = 1 \) under default \( d_0 \), since a DM cannot choose an alternative not in the attention set \( \Gamma(\cdot) \). Given the assumption that all DMs also pay attention to the default option, by choosing \( y = 1 \) a DM reveals that she prefers \( y \) to \( x \). Thus, when inattention drives default effects and individuals always pay attention to the default option, the assumptions underlying our reduced-form approach will obtain.\(^{114}\)

In related work on stochastic choice, Manzini and Mariotti (2014) make an assumption analogous to Equation (3.33) to model DMs who (with some probability) may or may not consider options besides the default.

In this model, variation in whether decision-makers are consistent is governed by variation in \( \Gamma_i \). The relationship between preferences and consistency is thus governed by the relationship between \( u_i \) and \( \Gamma_i \). In the next two sub-sections, we consider two alternative possibilities for what might drive variation in \( \Gamma_i \) and what each implies for the relationship between consistency and preferences.

### C.1.2 Heterogeneous Attention Costs

Here we consider a model in which paying attention to the non-default option requires the DM to incur some utility cost \( k_i \geq 0 \). We assume decision-makers fail to consider the non-default option whenever the cost of paying attention to the non-default option, \( k_i \), exceeds a threshold value \( \bar{k} \). For example, \( k_i \) may reflect the decision-maker’s cognitive ability, other demands on his or her attention, or prior experience with the choice being made. A decision-maker is consistent, \( k_i = 1 \), if and only if \( k_i < \bar{k} \).

\(^{114}\)One could imagine extending this approach to incorporate additional data. Note that, like frame monotonicity, property (3.33) could be tested if one is able to observe an individual’s choice across multiple frames, or if one could observe attention directly (such as by interviewing decision-makers after their choice or by employing an eye scanner). In addition, note that when we move beyond the binary case, assumption (3.33) would justify the assumption that active choices reveal a preference for the chosen option over the default option, but not the stronger assumption that active choices reveal preferences over the entire menu.
Assuming that $k_i$ is distributed in the population with a cumulative distribution function $G(.)$, we will have that $E[c_i] = G(\bar{k})$. Whether decision-quality independence holds in this model depends on the empirical correlation between the determinants of decision-makers’ attention costs, $k_i$, and their preferences, represented by $\bar{u}_i$. In particular, we will have $\text{cov}(c_i, y^*_i) = 0 \iff p(k_i < \bar{k}; \bar{u}_i > 0) = p(k_i < \bar{k}) p(\bar{u}_i > 0)$. Thus a sufficient condition for decision-quality independence is if $k_i$ is distributed independently of $\bar{u}_i$.

When $k_i$ and $\bar{u}_i$ are correlated, this correlation should be taken into account to estimate the distribution of preferences in the population. There are two empirical strategies one might employ. The first is to collect data on variables likely to be highly correlated with $k_i$ and implement a matching-on-observables approach. The second strategy would be to use decision-quality instruments to exogenously increase $k_i$. For example, providing decision-makers with practice making similar decisions or manipulating the intensity of the default effect are natural candidates for decision-quality instruments.

With variation in a decision-quality instrument, this model begins to resemble the one in Section 3.3.2 of the paper. With a joint normal distribution of $\bar{u}_i$ and $\log(k_i)$ and a homogeneous effect of a change in the decision-quality instrument $z$ on $\log(k_i)$, this becomes identical to the latent variable model outlined in that section, so we can trace out the propensity to optimize as a function of the fraction of optimizers at a given level of $z$, and extrapolate to recover the full distribution of $\bar{u}_i$ and $k_i$. One can naturally imagine similar models relying on weaker functional form assumptions.

### C.1.3 Stakes-Based Attentiveness

In this model, the DM decides whether to pay attention by comparing the cost of doing so against the benefit they stand to gain. In making this decision, the agent knows the absolute value of the difference in utility between the available options, but not which option has the higher utility. That is, agents know the utility amount at stake, but not which option they prefer. For example, consider an employee selecting a retirement savings plan. The
employee may know how much selecting the right retirement savings plan matters to her and how costly it is to learn about the menu of plans, but she may not actually know which plan is best for her without incurring utility costs from making the comparison.

Assume the DM knows the utility “stakes” of the meta-decision of whether to consider the non-default option, $|\bar{u}_i|$, and must decide whether to incur the cost of attention, $k_i$. As above, we assume that individuals who pay the cost select their most-preferred option under both defaults whereas individuals who do not simply select the default.

Suppose that decision-makers believe (ex-ante) that $y = 0$ is best with probability $\omega_i$. Consider the DM’s problem when the default is $d_0$. If she pays attention, with probability $\omega$ she will end up staying with $y = 0$ and with probability $1 - \omega$ she will discover that she prefers $y = 1$ (i.e., that $\bar{u} > 0$) and pick $y = 1$. Whichever option she prefers, she will incur the attention cost $C$. If she doesn’t pay attention, she will end up with $y = 0$ with certainty. Thus the net utility gain to paying attention when the default is $d_0$ is given by $(1 - \omega) |\bar{u}| - C$. The decision-maker will choose to pay attention to both options under the following condition:

$$\Gamma_i(X, d_0) = X \iff (1 - \omega_i) |\bar{u}_i| - k_i > 0$$

Similarly, when the default is $d_1$, the DM will choose to consider the non-default option when

$$\Gamma_i(X, d_1) = X \iff \omega_i |\bar{u}_i| - k_i > 0$$

Thus for an agent to choose to consider both options under both frames, it must be the case that

$$\Gamma_i(X, d_0) = \Gamma_i(X, d_1) = X \implies \frac{k_i}{|\bar{u}_i|} < \omega < 1 - \frac{k_i}{|\bar{u}_i|}$$

Note that the condition is guaranteed to fail when $k_i \geq |\bar{u}_i|$ (since $\omega \in [0, 1]$), i.e. when

If agents could not account for the amount at stake, this model would reduce to the previous one in which heterogeneity in attentiveness depended solely on individual characteristics. If agents knew the precise utility gain that they would achieve by considering the non-default option, the model would become formally identical to the costly opt-out model considered below.
the cost of attentiveness exceeds the potential benefits. Note that we can also write the above condition for attentiveness in both frames as \( k_i < \min \left\{ (1 - \omega_i) |\bar{u}_i|, \omega_i |\bar{u}_i| \right\} \), which highlights that a DM with sufficiently low \( k_i \) will always consider both options.

Because this model is a special case of the general inattention model described above, we know that frame monotonicity and the consistency principle will be satisfied. Note though that although DMs who are fully attentive under both defaults will be consistent, the converse is not true. For example, a DM who prefers \( y = 1 \) and has \( \omega < \min \left\{ \frac{|\bar{u}|}{c}, 1 - \frac{|\bar{u}|}{c} \right\} \) will choose \( y = 1 \) consistently, even though she follows the default under \( d_1 \) and only considers both options under \( d_0 \).

In general, decision-quality independence will not hold in this setting. Even when attention costs \( k_i \) are orthogonal to preferences, \( |\bar{u}_i| \) may nonetheless be correlated with both. That is, individuals with high \( |\bar{u}_i| \) will be more likely to pay attention to both options (and thus to choose consistently) and also may be more likely to prefer one option to the other (i.e., preference intensity may be correlated with ordinal preferences). However, conditional decision-quality independence may still hold when one can observe sufficient characteristics to control for both \( k_i \) and \( |\bar{u}_i| \), in which case matching on observables will allow us to recover population preferences. That is, under stakes-based attention models, the observer should control for variation among decision-makers associated with the costs of attention and the perceived utility stakes in the underlying decision. For example, in the retirement savings plan context, one could solicit and control for 1) the individual’s knowledge of the definitions of various aspects of retirement plans, and 2) for the self-reported importance of the savings decision. Similarly, valid decision-quality instruments will consist of variation in the choice environment that affects the cost of attention or the perceived stakes of the decision monotonically for all individuals. For example, a researcher might emphasize the importance of the decision to some participants in an experimental intervention.
C.2 Costly Opt-Outs

In this model, we assume a DM can incur a perceived utility cost $\gamma_i \geq 0$ in order to choose an option that is not the default. Here we assume that the utility cost is neoclassical, such as an administrative fee for opting into an alternative retirement plan. Although we initially focus on the case in which $\gamma_i$ reflects a real cost (i.e., the perceived cost equals the true cost), this formal model also captures “as if” transaction costs that nonetheless shape decision-makers’ behavior, as explored in Appendix C.2 below. A defining feature of this class of model is that the decision-maker knows the potential utility gain from choosing the non-default option and selects it if and only if the benefit from doing so exceeds the perceived opt-out cost. DM’s choice is thus given by

$$y_i(d) = \arg \max_{y \in X} u_i(y) - \gamma_i \{c_i \neq d\}$$

When $d_i = d_0$, the solution to this problem is given by $y_{i0} = 1 \iff \bar{u}_i > \gamma_i$. When $d = d_1$, the solution is given by $y_{i1} = 1 \iff -\bar{u}_i < \gamma_i$. We can summarize the three distinct possibilities for the choices of individual $i$ as follows:

$$y_{i0}, y_{i1} = \begin{cases} (0, 0) & \text{if } -\bar{u}_i > \gamma_i \\ (0, 1) & \text{if } -\bar{u}_i < \gamma_i, \bar{u}_i < \gamma_i \\ (1, 0) & \text{if } \bar{u}_i > \gamma_i \end{cases}$$

From (3.34), it is straightforward to verify that the consistency principle and frame monotonicity will hold. The two statistics studied in our paper will be given in this model by

$$y_i^* = 1 \iff \bar{u}_i > 0$$

$$c_i = 1 \iff |\bar{u}_i| > \gamma_i$$

When transaction costs are homogenous, $\gamma_i = \gamma \forall i$, the average (ordinal) preferences of the consistent decision-makers is given by: $E[y_i^* | c_i = 1] = P(\bar{\pi}_i > 0 | \bar{\pi}_i \in (-\infty, -\gamma] \cup [\gamma, \infty))$,}

141
or

$$E[y^*_i | c_i = 1] = \frac{1 - F(\gamma)}{1 - F(\gamma) + F(-\gamma)}$$

Similarly, for the inconsistent decision-makers we have

$$E[y^*_i | c_i = 0] = P(\bar{u}_i > 0 | \bar{u}_i \in (-\gamma, \gamma)),$$

or

$$E[y^*_i | c_i = 0] = \frac{F(\gamma) - F(0)}{F(\gamma) - F(-\gamma)}$$

Note that heterogeneity in decision-makers’ consistency in this model is driven by heterogeneity in the intensity of their preferences as well as the size of their transaction costs. Consequently, decision-quality independence will not generally be satisfied:

$$\text{cov}(y^*_i, c_i) = p(\bar{u}_i > \gamma) - p(\bar{u} > 0) \cdot p(\bar{u}_i < -\gamma \text{ or } \bar{u}_i > \gamma)$$

which will equal zero if and only if the distribution of preferences happens to satisfy

$$p(\bar{u}_i > \gamma | \bar{u}_i > 0) = p(\bar{u}_i < -\gamma | \bar{u}_i < 0).$$

That decision-quality independence usually fails here is not surprising: whether an individual is consistent in this model depends strongly on her preferences.

Nevertheless, additional structure can make the statistics on ordinal preferences studied in the body of the paper sufficient for optimal policy. One can show that when 1) $\gamma$ is homogeneous and 2) the distribution of $\bar{u}_i$ is single-peaked and symmetric, that the optimal default for a utilitarian social planner is $d = 1$ if and only if $E[y^*_i | c_i = 1] > \frac{1}{2}$.\textsuperscript{117} Moreover, these sufficient conditions for $E[y^*_i | c_i = 1] > \frac{1}{2}$ to determine the optimal policy obtain regardless of whether $\gamma$ is a real utility cost that enters the planner’s objective or merely a normatively irrelevant “as-if” cost.

Decision-quality instruments are immensely useful in this model, both for accounting for selection into the consistent subgroup and for identifying preference intensity. In this model changes in the cost of opting out constitute valid decision-quality instruments. Re-

\textsuperscript{116} $\text{cov}(y^*_i, c_i) = E[y^*_i c_{ii}] - E[y^*_i]E[c_{ii}] = p(y^*_i = c_{ii} = 1) - p(y^*_i = 1) \cdot p(c_{ii} = 1)$.

\textsuperscript{117} We conjecture that this statement also obtains when $\gamma$ is heterogeneous and independent of $\bar{u}_i$. Formal proofs of these statements, along with considerations of more general cases, will be contained in a forthcoming paper.
ductions in these costs could be obtained, for example, by easing the administrative requirements (such as paperwork) for choosing the non-default option. Suppose that transactions costs change from $\gamma_i$ to $\gamma'_i \leq \gamma_i$, with $\gamma'_i < \gamma_i$ for some $i$. Then variation in transactions costs will meet the criteria for being a decision-quality instrument and we will have:

$$
(y^*_i(\gamma), y^*_i(\gamma'), y_{ii}(\gamma'), y_{ii}(\gamma)) =
\begin{cases}
(0, 0, 0, 0) & \text{if } \bar{u}_i < -\gamma_i \\
(0, 0, 0, 1) & \text{if } \bar{u}_i \in [-\gamma_i, -\gamma'_i] \\
(0, 0, 1, 1) & \text{if } \bar{u}_i \in [\gamma'_i, \gamma_i] \\
(1, 1, 1, 1) & \text{if } \bar{u}_i > \gamma_i
\end{cases}
$$

The second and fourth cases correspond to the sometimes-consistent decision-makers whose ordinal preferences are captured by the statistic $Y_S$ in Section 3.3. Figure 3.3 depicts the different cases in Equation (3.35), given two values of a decision-quality instrument.

Because of the two-sided, symmetric nature of selection into the consistent subgroup in this model, we can identify the cardinal utility parameters governing the distribution of $\bar{u}_i$ and $\gamma_i$. With sufficient (observable) variation in $\gamma_i$ and/or functional form assumptions on the joint distribution of $\gamma_i$ and $u_i$, one can back out the underlying structural parameters using maximum-likelihood estimation or semi-parametric techniques. The setup of these
estimation strategies is similar to that of the one in Section 3.3.2, except that selection is two-sided here and one-sided in Section 3.3.2.

Here we will focus on the situation where $\gamma$ is homogeneous and known. Figure 3.4 illustrates the recovery of preference intensity given arbitrarily rich (exogenous) variation in $\gamma$. The top panel depicts Equation (3.34) for some $\gamma$. From this panel we can see that Equation (3.34) implies that given some $\gamma$, $E[y_{i0}(\gamma)]$ tells us about the fraction of consistent choosers with $\bar{u}_i > \gamma$, while $E[y_{i1}(\gamma)]$ tells us about the fraction of consistent choosers with $\bar{u}_i < -\gamma$. The bottom panel shows how given rich variation in $\gamma$, so that we know $E[y_{i0}(\gamma)]$ and $E[y_{i1}(\gamma)]$ as a function of $\gamma \in [0, \infty)$, we can recover the full cumulative distribution function of $\bar{u}_i$, where the units of $\bar{u}_i$ are measured in the same units as $\gamma$.

**C.3 Variants of the Costly Opt-Out Model**

This section considers models in which default sensitivity arises for reasons apart from transaction costs but that can be formally modeled along the same lines. An important motivation for these models is the fact that default sensitivity is observed among decision-makers even when the stakes are large and opt-out is inexpensive (Bernheim, Fradkin and Popov, 2014; Carroll et al., 2009; Chetty et al., 2014), which suggests that pure neoclassical transaction costs are insufficient to explain behavior. Note that although the positive behavioral models are similar, the distinction between misperceived and real transaction costs does matter for setting the optimal policy, as misperceived costs should not be incorporated into the social welfare function (i.e. they should be excluded from the $\gamma$ parameter in Appendix B).

---

118 We defer to future work situations in which $\gamma$ is heterogeneous and may have units that are difficult to quantify. Identifying a model with heterogeneous $\gamma$ might utilize exogenous variation in $\bar{u}$. This variation could come from varying the relative price of $y = 0$ and $y = 1$, for example, but allowing for such variation leads to a considerably more sophisticated model than the binary choice model we employ here.
observed behavior:

<table>
<thead>
<tr>
<th>$y_0$</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3.4: Identifying Preference Intensity with a Decision-Quality Instrument
Procrastination

The model in this section shows how present bias can contribute to default sensitivity by inflating perceived opt-out costs, which must be presently-incurred, relative to a benefit that is realized in the future (Laibson, 1997). The model is a simplified version of the one presented in Carroll et al. (2009), which employs non-binary menus and a richer dynamic structure.

There are two time periods in the model, $t = 1, 2$. At $t = 1$, decision-makers choose between $y = 0$ and $y = 1$. As above, considering the non-default option requires incurring some opt-out cost of $c_i$, for example, the disutility of expending mental energy for the decision-maker to figure out which option she prefers. At $t = 2$, the decision-maker receives the option she selected at $t = 1$ and realizes the associated utility. The DM’s time preference is captured by $\delta_i \in (0, 1]$. Additionally, the DM may also be present-biased, in that she perceives the utility consequences of current period decisions to be greater than future ones (over and above her discount rate $\delta_i$). We denote the DM’s present bias by $\beta_i \in (0, 1]$. Thus, for a DM who prefers $y = 1$ facing a default of $d_0$ in $t = 1$, the (perceived) net utility effect of selecting the non-default option is given by

$$U_i(1, d_0) = \beta_i \delta_i u_i - c_i$$

Let $\gamma_i \equiv \frac{\delta_i}{\beta_i}$ denote the current value of true utility costs from the perspective of the second period, which will be equivalent to opt-out costs in the previous section (since they are valued in the same period that utility over the menu options is realized). In this model, a decision-maker selects the non-default option if and only if the benefit to doing so exceeds the opt-out costs from the present-biased perspective:

$$y_{i0} = 1 \iff \frac{\gamma_i}{\beta_i} > \frac{\gamma_i}{\beta_i}$$

(3.36)
\[ y_{i1} = 0 \iff \pi_i < -\frac{\gamma_i}{\beta_i} \]  

(3.37)

Comparing Equations (3.36) and (3.37) to Equation (3.34) reveals that the procrastination model is formally equivalent to the standard costly opt-out model if we re-define the “opt-out cost” to be the true current value of opt-out costs inflated by the degree of the DM’s present-bias. Thus the key question for identifying the inconsistent choosers’ preferences is understanding the empirical correlation between \( \pi_i \) and \( \frac{\gamma_i}{\beta_i} \) in the population of decision-makers. From the perspective of social welfare, the only difference between this model and the previous one is that when we adopt the long-run view of welfare, only a fraction of the as-if opt-out costs are normatively relevant.

**Status-Quo Bias**

One candidate explanation for default effects is that decision-makers are more likely to choose whichever option represents a continuation of the status quo and perceive the default to be such a continuation. In this model, the frames denote which option represents a continuation of the status quo and which represents a change. A straightforward way to model status quo bias is to assume that individuals incur a utility cost when deviating from the status quo.\(^{119}\) That is, if the status quo is for an individual to end up with \( y = 0 \), she will only select \( y = 1 \) when the utility gains to doing so exceed a positive threshold. The only difference between this model and the one in Appendix C.2 is that the utility loss associated with deviating from the status quo is non-neoclassical. However, such behavior can be modeled in exactly the same way as default sensitivity that arises in response to a neoclassical transaction cost. Good candidates for decision-quality instruments would be variations in the *extent* to which an option is framed as a continuation of the status quo, such as by emphasizing or downplaying that aspect of the option at the time of decision-making.

\(^{119}\)Researchers and psychologists disagree about whether a status quo bias is normatively relevant or not – that is, whether status quo effects arise from non-standard preferences as opposed to an optimization failure.
Anchoring

In this model designating some option as the default creates a psychological pull towards that option, due to an anchoring effect. Relative to utility over $x$ and $y$, DM $i$ chooses as if she values the psychological pull of the default option at $\delta_i$. DM’s decision utility of option $y \in X$ is given by

$$\hat{u}_i(y,d) \equiv u_i(y) + \delta_i 1\{y = d\}$$

where $1\{y = d\}$ indicates whether option $y$ is the default. We assume that an individual maximizes her decision utility, i.e. that $y_i(d) = \arg\max \hat{u}(y,d)$. When $d = d_0$, the solution to this problem is given by $y_{i0} = 1 \iff \bar{u}_i > \delta_i$. When $d = d_1$, the solution is given by $y_{i1} = 1 \iff \bar{u}_i > -\delta_i$. We can summarize the three distinct possibilities for the choices of individual $i$ as follows:

$$
\begin{align*}
(0,0) & \quad \text{if } \bar{u}_i < -\delta_i \\
(0,1) & \quad \text{if } -\delta_i < \bar{u}_i < \delta_i \\
(1,1) & \quad \text{if } \bar{u}_i > \delta_i
\end{align*}
$$

Equation (3.38) reveals that although the psychological motivation is quite different, this model too is formally equivalent to the costly opt-out models described above. For the binary case considered here, the difference between the two models reduces to whether the effect of the frame on decision-making is modeled as a cost (as above) or as a benefit (as here).\textsuperscript{120} Candidate decision-quality instruments might vary the psychological pull of the default option ($\delta$), for example by varying its prominence relative to the rest of the menu.

Rational Attention

Here we consider the possibility that the opt-out costs are the psychological costs of paying attention to the non-default option. When the individual rationally incurs these costs, the

\textsuperscript{120}Outside of the binary setting, Bernheim, Fradkin and Popov (2014) note that another way to distinguish between the models and the opt-out costs models is that anchoring should cause choices to cluster around the default, whereas opt-out costs should create a trough in the distribution of choices around the default.
model in Appendix C.1 becomes a model of rational inattention. Formally, we can suppose that the individual faced with default $d$ first chooses $\Gamma(d_0)$

$$d_0 : \max_{\Gamma(d_0) \in \{\{0\}, \{0, 1\}\}} u_i(y(\Gamma(d_0))) - \gamma \mathbb{1}\{1 \in \Gamma(d_0)\} \text{ s.t. } y(\Gamma(d_0)) = \arg \max_{y \in \Gamma(d_0)} u_i(y)$$ (3.39)

$$d_1 : \max_{\Gamma(d_1) \in \{\{1\}, \{0, 1\}\}} u_i(y(\Gamma(d_1))) - \gamma \mathbb{1}\{0 \in \Gamma(d_1)\} \text{ s.t. } y(\Gamma(d_1)) = \arg \max_{y \in \Gamma(d_1)} u_i(y)$$ (3.40)

The constraint in this optimal attention problem corresponds to the choice rule described by equation (3.32) in Appendix C.1, and the key assumption from that Section, that the individual always pays attention to the default option (Equation (3.33) before), is embedded in the possible choices of $\Gamma(\cdot)$. When we endogenize attention in this way, the inattention model becomes exactly identical to the opt-out costs model, as characterized by Equation (3.34). This model is also a version of the planner-doer model of Fudenberg and Levine (2006), where the behavior of the “doer” is given by Equation (3.32) and the behavior of the “planner” – who regulates the behavior of the doer at some cost – by Equations (3.39) and (3.40).

### C.4 Defaults as Advice

Here we present a simple model based on the idea that decision-makers who lack information about the utility of the available options may interpret the default as advice from the planner about which option is likely to be in their best interest. That is, decision-makers have some prior beliefs about the utility they would derive from either choice and they update that prior based on a signal in the form of the default. The model we develop here is a simplified version of that of Caplin and Martin (2012). Although this model is a plausible description of behavior, it poses normative problems for the framework we have set out.

---

121 As with many models of rational inattention, this model is subject to an infinite regress critique (Conlisk, 1996). One might wonder, for instance, how the individual allocates attention to the problem of allocating attention among menu options. We assume the individual perfectly understands the costs and benefits of attention allocation, but acknowledge that this does not fully overcome the conceptual difficulty.
With respect to decision-makers’ ex ante preferences, the model violates frame separability because which option is the default affects which option decision-makers expect to have higher utility. Likewise, with respect to ex post preferences (i.e., which option turns out to be better for decision-makers after all uncertainty is resolved), the consistency principle may be violated; decision-makers who have a prior that \( y = 0 \) conveys much higher utility than \( y = 1 \) may select \( y = 0 \) under both defaults but nonetheless turn out to prefer \( y = 1 \).

In this section, we show that despite these problems, the framework described in the text, especially those related to identifying consistent choosers (Propositions 1 and 4), can still provide interesting information in this context. In particular, understanding the determinants of consistency will shed light on the signal value of defaults, and observing choices of consistent individuals will convey information on the distribution of priors beliefs about the relative utility of the two options.

Formally, denote individual \( i \)'s expected utility of \( y = 1 \) relative to \( y = 0 \) by \( E[\bar{u}_i] = E[u_i(1) - u_i(0)] \). Upon observing the default, the individual updates this expected utility to \( E[\bar{u}_i|d] \) and chooses \( y = 1 \) iff \( E[\bar{u}_i|d] > 0 \). For simplicity, we will assume the change in expected relative utility upon updating is symmetric across the two defaults and denoted by
\[
v_i \geq 0;  \text{ }^{122}
\]
\[
E[\bar{u}_i|d_1] = E[\bar{u}_i] + v_i
\]  \hspace{1cm} (3.41)
\[
E[\bar{u}_i|d_0] = E[\bar{u}_i] - v_i
\]  \hspace{1cm} (3.42)

Frame monotonicity will be satisfied in this model, since \(E[\bar{u}_i|d_1] \geq E[\bar{u}_i|d_0]\). \text{ }^{123} In fact, this model is quite similar to the model of costly opt-out in Appendix C.2, with a different interpretation of the underlying parameters. Combining (3.41) and (3.42) yields

\[
(y_{i0}, y_{i1}) = \begin{cases} 
(0, 0) & \text{if } -E[\bar{u}_i] > v_i \\
(0, 1) & \text{if } -E[\bar{u}_i] < v_i, \; E[\bar{u}_i] < v_i \\
(1, 1) & \text{if } E[\bar{u}_i] > v_i
\end{cases} \hspace{1cm} (3.43)
\]

which closely resembles (3.34). The key difference between this model and the costly opt-out one is that instead of true (full information) preferences (\(\bar{u}\)) and cognitive costs (\(\gamma\)), our methods will identify information about prior beliefs (\(E[\bar{u}_i]\)) and the signal value of defaults (\(v_i\)). These insights are obtained once again by comparing consistent and inconsistent

---

\text{122} We can use a slightly different set-up to develop the same intuition and make the updating upon learning the default more explicit. Suppose each DM \(i\) has prior belief \(P(\bar{u}_i > 0) = \alpha_i\), and believes the default is correct with probability \(\beta_i > \frac{1}{2}\), regardless of whether the signal is \(d_0\) or \(d_1\) (which assumes some symmetry like in the other model considered here). The signal she receives is independent of her prior. An individual updates her subjective probability that \(y\) is preferred to \(x\) as follows:

\[
P(\bar{u}_i > 0|d_1) = \frac{P(\bar{u}_i > 0, d_1)}{P(\bar{u}_i > 0, d_1) + P(\bar{u}_i < 0, d_1)} = \frac{\alpha_i \beta_i}{\alpha_i \beta_i + (1 - \alpha_i)(1 - \beta_i)}
\]

\[
P(\bar{u}_i > 0|d_0) = \frac{P(\bar{u}_i > 0, d_0)}{P(\bar{u}_i > 0, d_0) + P(\bar{u}_i < 0, d_0)} = \frac{\alpha_i(1 - \beta_i)}{\alpha_i(1 - \beta_i) + (1 - \alpha_i) \beta_i}
\]

Suppose that the individual chooses \(y = 1\) iff \(p(\bar{u}_i > 0|d) > \frac{1}{2}\), which would be true if the subjective distribution of \(\bar{u}_i\) were symmetric for all \(i\). Then the analogue to equation (3.43) is as follows:

\[
(y_{i0}, y_{i1}) = \begin{cases} 
(0, 0) & \text{if } \alpha < 1 - \beta_i \\
(0, 1) & \text{if } 1 - \beta_i < \alpha < \beta_i \\
(1, 1) & \text{if } \alpha > \beta_i
\end{cases}
\]

which conveys a similar intuition about using our approach to learn about prior beliefs (\(\alpha_i\) in this model) and the signal value of defaults (\(\beta_i\) in this model).

\text{123} Note that this obtains even without the symmetry assumption employed above, so long as \(E[\bar{u}]\) is shifted (weakly) upward by \(d_1\) and downward by \(d_0\). Frame monotonicity fails when some individual perceives the default-setter as providing bad advice, i.e. \(v_i < 0\) for some \(i\).
choosers, as consistent choosers will choose the option favored by their prior beliefs regardless of the frame. One can imagine adding interventions analogous to decision-quality instruments here by attempting to manipulate the perceived signal value of the default option. For example, one could provide some decision-makers with a salient disclaimer that the default is not intended to convey advice.

Because of this key difference in underlying parameters and the violation of frame separability in this model, our methods will not be useful for setting optimal frames when defaults are interpreted by individuals as advice. But this should not be surprising: when individuals assume that the agent setting the default has preference-relevant information that they do not have, it should not be possible to infer their preferences directly from their choices conditional on alternative default policies. Indeed, absent information about preferences that does not come from observing choices under different defaults, there is a striking recursion embedded in the optimal framing problem for this model: the planner’s optimal choice of default must depend on inferring preference information from choices, but default effects only occur in the first place because individuals assume the planner already has some preference information.

**D. Generalizations**

**D.1 Non-Binary Frames, Varying Intensity**

Consider choice situations in which an individual $i$ chooses from a binary menu $X = \{0, 1\}$ under one of multiple frames that vary in their intensity, $d \in \{d_0, d_1, \ldots, d_J\}$. For example, if $y = 0$ has greater up-front costs than $y = 1$, the frames might describe the extent to which those costs are made salient to the decision-maker, with $d_1$ denoting the frame in which the costs are least salient and $d_J$ denoting the frame in which they are most salient. Alternatively, the decision could be one in which the decision-maker must choose whether to purchase a good for a given price, and the frame describes the reference point with which
the decision-maker has been presented (if decision-makers are subject to an anchoring effect, the larger the reference point the more they might be willing to pay).

The key assumption that will let us apply the tools from the rest of the paper to this setting is that the frames can be ordered according to their intensity:

\[ y_i(d_j) \geq y_i(d_{j'}) \quad \forall i, j \succ j' \quad (3.44) \]

where the ordering of frames is without loss of generality. In words, (3.44) requires that if a decision-maker chooses an option under one frame, he or she will also choose that option under any frame that pushes more intensely in that option’s direction. With (3.44), frame monotonicity applies with respect to any two frames.

We also assume a global consistency principle:

\[ y_i(d_j) = y_i(d_{j'}) \quad \forall j, j' \Rightarrow y_i(d_j) = y_i^* \quad (3.45) \]

This is a natural extension of the consistency principle discussed earlier: if a decision-maker would select the same option in each frame, we assume that choice is her preferred one.

When (3.44) and (3.45) hold, the multi-frame setting can be reduced to the binary one studied in the rest of the paper. In particular, one can apply Proposition 1 after setting \( d_0 = d_0 \) and \( d_1 = d_J \) to obtain information about the set of decision-makers consistent with respect to all of the observed frames. Additionally, even when (3.45) does not hold, one can apply the matching estimator (Corollary 4.1) to recover information about the characteristics of consistent and inconsistent decision-makers at any two observed frames so that the researcher can investigate heterogeneity in which decision-makers are susceptible to low-intensity framing effect as opposed to both high- and low-intensity ones.

Additional structure beyond the global consistency principle allows one to recover even more information on preferences. To illustrate this, we note that this problem has an interesting relationship to the model of decision-quality instruments presented in Section
3.3. We alluded in Section 3.3 to the idea that valid decision-quality instruments can include those varying the intensity of a given framing effect. Formally, let \( d_1 = (z_l, d_0), \ d_2 = (z_h, d_0), \ d_3 = (z_h, d_1), \) and \( d_4 = (z_l, d_1) \) be the four possible frames. The consistency principle from Section 3.3, at \( z_l \), is equivalent to the global consistency principle (3.45). The idea in the previous paragraph, using the two most extreme frames, is exactly analogous to recovering the preferences of the consistent group at \( z_l \): \( Y_C(z_l) = E[y^*_i | c il = 1] \). The consistency principle at \( z_h \) in Section 3.3 also implies a second condition for consistency across \( d_2 \) and \( d_3 \), which will imply that \( Y_C(z_h) = E[y^*_i | c ih = 1] \). Frame separability in this model implies decision quality exclusion, and frame monotonicity implies decision quality monotonicity, so that all changes in behavior between \( d_1 \) and \( d_2 \), and between \( d_3 \) and \( d_4 \) tell us about the preferences of individuals consistent across \( (d_1, d_4) \) but not across \( (d_2, d_3) \), which allows us to recover the preferences of decision-makers whose choices depend on the intensity of framing, \( Y_S = E[y^*_i | c ih > c il] \).

**D.2 Multi-Dimensional Frames**

This sub-section considers choice settings that differ along multiple dimensions so that frames cannot be ordered by intensity. Consider choice situations in which an individual \( i \) chooses from a binary menu \( X = \{x, y\} \) under a frame vector \( d = (d_1, \ldots, d_J) \), so that each component of \( d \) encodes some feature of the choice environment. We assume each frame component \( d_j \) of \( d \) is discrete with two possible realizations: \( d_j \in \{d_{0j}, d_{1j}\} \).\(^{124}\) For example, a decision-maker’s choice between two options might be affected both by which option is presented first and by which option is framed as the default. In this example, \( d_1 \) could describe the order of the options and \( d_2 \) could describe which option is the default.

As before, denote choices under frame \( d \) by \( y_i(d) \). We will assume component-wise frame monotonicity:

---

\(^{124}\)It is straightforward to combine this approach with the one in the previous section, extending this setup to settings in which each frame component has multiple possible realizations that vary in their intensity.
where $d_{-j}$ is the vector consisting of all frame components other than $j$. Equation (3.46) implies that frame monotonicity holds for each component of the frame vector when all components are held fixed. It also requires that the direction of the effect of any one decision characteristic on choice be independent of other decision characteristics. For example, it must not be the case that making $y = 0$ the default induces more decision-makers to select $y = 0$ when 0 is listed first but that making 0 the default induces more decision-makers to select $y = 1$ when 0 is listed second.

As above, we assume a global consistency principle:

$$y_i(d) = y_i(d') \quad \forall d, d' \implies y_i(d) = y_i^*$$  

(3.47)

This principle means that whenever the individual would choose the same option in every frame, she prefers the option that she chooses.

When (3.46) and (3.47) hold, we can proceed similarly to the previous subsection using data on the two most extreme frames. To do this we can apply Proposition 1 with $d_0 = (d_{01}, d_{02}, ..., d_{0J})$ and $d_1 = (d_{11}, d_{12}, ..., d_{1J})$ to recover the preferences of globally consistent decision-makers, those whose choice is insensitive to all framing effects. We use this approach in the empirical illustration in Appendix . Further possibilities are generated by considering what additional structure might allow us to learn about preferences from the behavior of individuals who are consistent with respect to all frame components but one, all frame components but two, and so on.

**D.2.1 Application to Privacy Controls**

In this section we re-examine data collected by Johnson, Bellman and Lohse (2002), whose work inspired our running example of online privacy controls. In this study, the authors initially asked participants to complete an online survey about their health. Participants
were then asked whether they would like to receive additional surveys.\textsuperscript{125} There were two sources of framing effects for these solicited preferences. First the question was either framed negatively ("Do NOT notify me about more health surveys") or positively ("Notify me about more health surveys."). Second, the answers to the question were pre-selected to be either yes or no, so that the consumer would have to actively change their answer to avoid the default. We therefore employ the extension of our approach to multi-dimensional framing effects considered in Appendix D.2. In particular, $y$ indicates whether the individual gave permission to be contacted for additional surveys, $d_0$ is the situation in which the question was framed positively and the default was non-participation ("opt-in"), and $d_1$ is the situation in which the question was framed as a loss and the default was participation ("opt-out"). We do not use data from the other two frame combinations. This exclusion drops 139 of the sample of 277 individuals in the original study, leaving 138 observations. Although we do have some demographic variables, our sample size makes results from dividing the population into demographic groups imprecise.

Table 5 presents the results from this application. Several interesting patterns emerge. First, framing affects the participation decisions of almost half (48.2 percent) of the survey-takers; only 51.8 percent of the population are globally consistent, meaning their choice depends neither on the positive or negative wording of the question nor on the default, pre-selected answer. Second, of the globally consistent subgroup, over 90 percent prefer to participate. The size of this estimate is attributable to the fact that, under $d_1$, very few individuals opt out of participation, while under $d_0$, many more individuals opt in. Given that a very high fraction of globally consistent individuals participate in the surveys, we conclude that under decision-quality independence, an opt-out policy would maximize the fraction of participants ending up with their preferred option.

\textsuperscript{125}We focus on the data reported in “Study 1” of Johnson, Bellman and Lohse (2002).
Table 5: Participation in Online Surveys

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participation rate under $d_1$</td>
<td>0.9563</td>
</tr>
<tr>
<td></td>
<td>(0.0251)</td>
</tr>
<tr>
<td>Participation rate under $d_0$</td>
<td>0.482</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
</tr>
<tr>
<td>Fraction consistent</td>
<td>0.518</td>
</tr>
<tr>
<td></td>
<td>(0.070)</td>
</tr>
<tr>
<td>Fraction of consistent who prefer participation</td>
<td>0.929</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
</tr>
</tbody>
</table>

Note: standard errors are estimated using the delta method (see Appendix E).

D.3 Non-Binary Menus with Ordered Options

This section develops an approach for preference recovery over non-binary menus. There are many interesting possibilities for generalizations, but we focus here on choice situations $g \in G$ consisting of a fixed, finite menu of $K$ ordered options $X = \{1, 2, \ldots, K\}$ and one of two frames, $d \in \{d_h, d_l\}$. Intuitively, one can think of $d_h$ and $d_l$ as a “high” frame and a “low” frame. For example, we might suppose that an individual chooses from a menu of insurance plans, ordered from low-cost, low-benefit plans to high-cost, high-benefit plans, and the frame either emphasizes or de-emphasizes the individual’s risk of developing a serious illness. We will assume that we observe each individual $i$ in exactly one frame, as before. Recall that in the binary case, the consistency principle and frame monotonicity imply that individuals who choose the “low” option in the “high” frame prefer the low option. We will use this same intuition to develop an identification strategy for the non-binary setting.

The preferences of agent $i$ are represented by choice function $y^*_i \in X$. We continue to assume frame separability (A1), so that $y^*_i$ does not depend on $d$.

We strengthen the frame monotonicity assumption as follows:

D1 (Frame monotonicity for many options) For all individuals, $y_i(d_h) \geq y_i(d_l)$.

This frame monotonicity assumption imposes an implicit ordering on the menu and as-
assumes that all individuals are pushed in the same direction by the frames. We also strengthen the consistency principle with the following assumption

**D2** (Partition-Consistency principle) For all individuals $i$ and options $k \in X$,

\[
\begin{align*}
y_i(d_l) \geq k & \implies y_i^* \geq k \\
y_i(d_h) \leq k & \implies y_i^* \leq k.
\end{align*}
\] (3.48)

The name of this assumption comes from the following: suppose that we partition the menu into $X' = \{J, J+1, \ldots, K\}$ and $X'' = X \setminus X'$, for some $J$ and $K \geq J$. If the individual consistently chooses within $X'$ across both frames, so $y_i(d_h) \in X'$, and $y_i(d_l) \in X'$, then assumption (3.48) implies that $y_i^* \in X'$. Note also that the partition consistency principle implies the consistency principle used in previous sections: if $y_i(d_h) = y_i(d_l)$, then assumption (3.48) implies that $y_i(d) = y_i^*$. Finally, note that the partition consistency principle and frame monotonicity together imply that $\forall i$, $y_i(d_h) \geq y_i^* \geq y_i(d_l)$.

For each $k = 1, \ldots, K$, we define **partition consistency at $k$, $c^k_i$**, as follows

\[
c^k_i \equiv 1\{y_i(d_h) \leq k \text{ and } y_i(d_l) \leq k\} + 1\{y_i(d_h) > k \text{ and } y_i(d_l) > k\}
\]

Intuitively, $c^k_i$ captures whether an individual consistently chooses an option above or below $k$. Note also that frame monotonicity implies that one of the conditions inside each indicator function will be implied by the other condition.

**Proposition A2** Let $G_j(k) \equiv P(y_i(d_j) \leq k|d_i = d_j)$ for $k = 1, \ldots, N$, $j = h, l$ and let $G_j(0) \equiv 0$. Let $Y_k \equiv \frac{G_h(k)}{G_h(k) + 1 - G_l(k)}$ for $k = 0, \ldots, K$. Frame separability (A1), frame monotonicity (D1), partition consistency (D2), and unconfoundedness (A4) imply that for $k = 1, \ldots, K$,

(A2.1) The fraction of partition-consistent individuals at $k$ with $y_i^* \leq k$ is given by $P(y_i^* \leq k|c^k_i = 1) = Y_k$.

(A2.2) The fraction of partition-consistent individuals at $k$ is given by $E[c^k_i] = G_h(k) + 1 -$
The fraction of the population who prefer option $k$ is bounded as follows: $p(y_i^* = k) \in [G_i(k) - G_h(k - 1), G_h(k) - G_i(k - 1)]$.

If we additionally assume strong decision-quality independence, $\forall k, y_i^* \perp c_i^k$, then the fraction of the population who prefer option $k$ is $p(y_i^* = k) = Y_k - Y_{k-1}$.

Proof

Throughout the proof, we denote the fraction of individuals preferring some option $k$ by $\bar{\phi}_k \equiv p(y_i^* = k)$.

Proof of (A2.1) and (A2.2): Fix some $k \in \{1, ..., K - 1\}$. Let $X' = \{x_1, ..., x_k\}$ and $X'' = \{x_{k+1}, ..., x_K\}$ Note that we can write the many-choices problem into a binary menu choice problem between $X'$ and $X''$. Similarly, note that frame separability (A1), frame monotonicity (D1), partition consistency (D2), and partition unconfoundedness (A4) imply the binary analogues to these assumptions A1-A4. As such, (A2.1) and (A2.2) follows directly from the application of Proposition 1 to this problem.

Proof of (A2.3): First suppose that $k = 1$. Applying Proposition 2 to the binary menu choice problem with $X' = \{1\}$ and $X'' = \{2, ..., K\}$ implies that

$$E[\phi_1] \in [G_i(1), G_h(1)]$$

(3.49)

Note that this confirms the desired result for $k = 1$ since $G_h(0) = G_i(0) = 0$ by definition. Next, applying the same proposition for $k = 2$, we have $\bar{\phi}_1 + \bar{\phi}_2 \in [G_i(2), G_h(2)]$. Combined with (3.49), this implies

$$\bar{\phi}_2 \in [G_i(2) - G_h(1), G_h(2) - G_i(1)]$$

(3.50)

Similarly with $k = 3$, we have that $\bar{\phi}_1 + \bar{\phi}_2 + \bar{\phi}_3 \in [G_i(3), G_h(3)]$, and applying (3.49) and (3.50) implies that $\bar{\phi}_3 \in [G_i(3) - G_h(2), G_h(3) - G_i(2)]$. Proceeding recursively, suppose
that for some \( k \), we know that for any \( k' < k \),

\[
\overline{\phi}_{k'} \in [G_I(k') - G_h(k' - 1), G_h(k') - G_I(k' - 1)]
\]  

(3.51)

Then application of Proposition 2 to the binary menu choice problem with \( X' = \{x_1, \ldots, x_k\} \) yields \( \overline{\phi}_1 + \overline{\phi}_2 + \ldots + \overline{\phi}_k \in [G_I(k), G_h(k)] \), so \( \overline{\phi}_k \in [G_I(k) - (\overline{\phi}_1 + \overline{\phi}_2 + \ldots + \overline{\phi}_{k+1}), G_h(k) - (\overline{\phi}_1 + \overline{\phi}_2 + \ldots + \overline{\phi}_{k+1})] \). Applying the lower and upper bounds from (3.51) and simplifying yields the desired result.

**Proof of (A2.4):** Along with (A2.1), strong decision-quality independence implies that for any \( k \),

\[
P(y_i^* \leq k | c_i^k = 1) = P(y_i^* \leq k) = Y_k
\]  

(3.52)

Applying (3.52) at \( k = 1 \) yields

\[
\overline{\phi}_1 = Y_1
\]  

(3.53)

Applying (3.52) at \( k = 2 \) yields \( \overline{\phi}_1 + \overline{\phi}_2 = Y_2 \) and substituting equation (3.53) yields

\[
\overline{\phi}_2 = Y_2 - Y_1
\]

As in the proof of (A2.3), we proceed recursively to obtain the desired result. Given some \( k \), suppose that for any \( k' < k \) we have

\[
\overline{\phi}_{k'} = Y_{k'} - Y_{k'-1}
\]  

(3.54)

Applying (3.52) at \( k \) yields \( \overline{\phi}_1 + \overline{\phi}_2 + \ldots + \overline{\phi}_k = Y_k \). Applying (3.54) for \( \overline{\phi}_1, \ldots, \overline{\phi}_{k-1} \) and simplifying yields the desired result.

**Discussion of Proposition A2** If we partition the menu of choices into options above and below some option \( k \), then frame monotonicity and the partition-consistency principle transform the problem to a binary problem, allowing us to use earlier propositions to identify individuals whose preferred choice is above or below \( k \). The first two results, (A2.1)
and (A2.2), are therefore the analogue of Proposition 1 in this setting.

Return to the insurance example described above, where the frame either emphasizes or de-emphasizes the risk of serious illness. When some individuals choose a low-benefit, low-cost plan under the frame that emphasizes the risk of serious illness, our assumptions imply that they prefer an option with costs and benefits at least as low as the ones they choose. The first two results in Proposition A2 allow us to estimate the fraction of decision-makers who consistently choose an insurance plan that is above or below some specified cost-benefit level, and among those people, how many prefer the low-cost plan.

As in Proposition 2, we can also bound population preferences, reflected in (A2.3). In this case, the many-options problem has a new and interesting structure. Even if individuals are highly susceptible to framing effects when they prefer some option far away from $k$, our estimate for the fraction of people preferring option $k$ can still be precise, because the partition consistency principle permits us to ignore individuals who consistently choose options above or below $k$.

Finally, with a stronger version of the decision-quality independence assumption, we can recover the distribution of preferences for the full population. Strong decision-quality independence guarantees that the tendency to be partition-consistent for any partition is unrelated to an individuals’ preferences. Under strong decision-quality independence, obtaining the preferences of partition-consistent individuals will yield the distribution of preferred choices in the population. The equivalence of this problem to the binary problem implies that we could generalize other identification strategies from the binary case. For example, we can identify the preferences of the population using observables via a conditional strong decision-quality independence assumption (the generalization of the matching approach), and in the absence of any decision-quality independence assumptions we can study variation induced by a decision-quality instrument.

---

126This assumption implies our earlier definition of decision-quality independence, because individuals who are consistent across frames will be partition-consistent for all partitions. If we were to assume that individuals are partition-consistent only if they are fully consistent across frames, which is trivially true in the binary case, then the two assumptions about decision-quality independence would be equivalent.
E. Estimating Asymptotic Variance in Finite Sample

The body of the paper ignores finite sample concerns, but any empirical application, including the ones we undertake in Section 4, should account for finite sample concerns and report standard errors for the estimators we propose. This section derives estimators of asymptotic variance, which may be used to construct standard errors of the estimators of the population parameters of interest. Because all variables we work with in this paper are discrete and the data are presumed to come from independent random sampling of the population, our derivations here rely only on the characteristics of the binomial and multinomial distributions, along with the delta method. The standard errors we derive are incorporated into Stata program files that are available upon request from the authors.

Standard Errors for Proposition 1

To facilitate compact exposition, let us introduce some statistical notation. Note that $y_{i0}$ and $y_{i1}$ are two-valued random variables. Let $Y_0$ and $Y_1$ denote the population moments we wish to estimate from data as before. Let $\bar{y}_0$ and $\bar{y}_1$ denote the sample averages under $d_0$ and $d_1$, respectively. From the properties of the binomial distribution,

$$\sqrt{n} \begin{pmatrix} \bar{y}_0 - Y_0 \\ \bar{y}_1 - Y_1 \end{pmatrix} \stackrel{d}{\sim} N(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \frac{1}{\alpha_0} Y_0 (1 - Y_0) & 0 \\ 0 & \frac{1}{\alpha_1} Y_1 (1 - Y_1) \end{pmatrix}$$

and $\alpha_j = \frac{n_j}{n}$ for $j \in \{0, 1\}$, where $n_j$ denotes the number of individuals observed under frame $d_j$.

The first statistic from Proposition 1 is $Y_c$, which expressed as a function of the primitive parameters of the model is

$$Y_c = \frac{Y_0}{Y_0 + 1 - Y_1}$$  \hspace{1cm} (3.55)
which we estimate consistently via

\[ \hat{y}_c = \frac{\bar{y}_0}{\bar{y}_0 + 1 - \bar{y}_1} \]

Using the delta method, we know that \( V(\hat{y}_c) \simeq \frac{1}{n} \nabla Y_c \Sigma \nabla Y_c \), where \( \nabla Y_c \) is evaluated at \((Y_0, Y_1)\). Taking derivatives of (3.55) yields

\[ \nabla Y_c = \begin{pmatrix} \frac{1 - Y_1}{(Y_0 + 1 - Y_1)^2} \\ \frac{Y_0}{(Y_0 + 1 - Y_1)^2} \end{pmatrix} . \]

Simplifying the expression for \( V(\hat{y}_c) \) yields

\[ V(\hat{y}_c) \simeq \frac{(1 - Y_1)^2 Y_0(1 - Y_0)}{\alpha_0 (Y_0 + 1 - Y_1)^4 n} + \frac{Y_0^2 Y_1(1 - Y_1)}{\alpha_1 (Y_0 + 1 - Y_1)^4 n} \]

which, letting \( \bar{c} = Y_0 + 1 - Y_1 \) and using that \( V(\bar{y}_j) = \frac{Y_j(1 - Y_j)}{n_j} \) for \( j = 0, 1 \), simplifies to

\[ V(\hat{y}_c) = \left( \frac{1 - \hat{c}}{\bar{c}} \right)^2 V(\bar{y}_0) + \left( \frac{\hat{c}}{\bar{c}} \right)^2 V(\bar{y}_1) \]  

(3.56)

We can consistently estimate the variance of the asymptotic distribution of our estimator \( \hat{y}_c \) by replacing all the terms in (3.56) with the corresponding sample means, i.e. replacing \( Y_0 \) with \( \bar{y}_0 \) and \( Y_1 \) with \( \bar{y}_1 \).

Note also that we can estimate the variance of the asymptotic distribution of our estimator for \( \bar{\sigma} = E[c_i] \). Write the estimator itself as:

\[ \hat{\sigma} = \bar{y}_0 + 1 - \bar{y}_1 \]

and the variance of this estimator is simply:

\[ V(\hat{\sigma}) = V(\bar{y}_0) + V(\bar{y}_1) \]  

(3.57)
Standard Errors for Matching Estimator

The standard errors for the matching-on-observables estimator are considerably more complicated, due to the presence of multiple demographic groups and the use of weights that must themselves be estimated from data. We discuss two solutions here, one based on an estimation strategy imposing parametric restrictions for how observable characteristics are related to choices under either frame, and another adopting a fully non-parametric approach.

First, one can simply estimate our model via

\[ E[y_i|d, w] = f(w_i, \theta) + 1\{d_i = d_1\}g(w_i, \theta') \]

where \(f()\) and \(g()\) are specified up to vectors of parameters \(\theta\) and \(\theta'\), which are estimated from data. For example, we could implement a linear model with uni-dimensional \(w\):

\[ E[y_i|d, w] = \alpha + \beta w_i + 1\{d_i = d_1\}(\gamma + \delta w_i) \]

This equation can be estimated by a least squares linear probability model, and then the ingredients of the matching estimator are given by

\[ \hat{Y}_c(w) = \frac{\hat{\alpha} + \hat{\beta} w_i}{1 - \hat{\gamma} - \hat{\delta} w_i} \]

\[ \hat{E}[c_i|w] = 1 - \hat{\gamma} - \hat{\delta} w_i \]

\[ \hat{E}[c_i] = \frac{1}{n} \sum_i (1 - \hat{\gamma} - \hat{\delta} w_i) \]

\[ \hat{E}[y_i^*] = \frac{1}{n} \sum_i \hat{Y}_c(w) \]

\[ \hat{E}[y_i^*|c_i = 0] = \hat{Y}_c(w) = \frac{1}{n} \sum_i \frac{1 - \hat{E}[c_i|w = w_i]}{1 - \hat{E}[c_i]} \hat{Y}_c(w_i) \]

This estimation strategy can be implemented via straightforward post-regression estimation, and standard errors may be estimated using the a straightforward non-parametric
bootstrap.

When \( w \) is discrete, taking values \( w_1, \ldots, w_J \) one may derive straightforward delta-
method-based standard errors. We provide analytical formulae for the variance and the
gradients of parameters of interest, which may be straightforwardly incorporated into a
matrix-based programming language, such as MATLAB, to calculate the variance of the
estimators with discrete demographic groups. MATLAB code illustrating this procedure,
used for the illustration in the body of the paper, is also available upon request from the
authors.

The primitive parameters of the discrete-characteristics model are, for each \( w, Y_{0w}, Y_{1w}, \)
and \( p_w = p(w_i = w) \). When the \( w \)'s are non-stochastic, such as when the researcher wishes
to estimate preferences for a population with a known distribution of observable character-
istics, the last of these may be excluded; the resulting modification of the variance
estimation procedure below is straightforward. We denote the estimators of these quan-
tities by \( \bar{y}_0w, \bar{Y}_{1w}, \) and \( \hat{p}_w \). Now we construct the variance covariance matrix of the vec-
tor primitive parameters. Letting \( \theta = (Y_{0w_1}, Y_{0w_2}, \ldots, Y_{0w_J}, Y_{1w_1}, \ldots, Y_{1w_J}, p_{w_1}, \ldots, p_{w_J})', \) and
\( \hat{\theta} = (\bar{y}_{0w_1}, \bar{y}_{0w_2}, \ldots, \bar{y}_{0w_J}, \bar{y}_{1w_1}, \ldots, \bar{y}_{1w_J}, \hat{p}_{w_1}, \ldots, \hat{p}_{w_J}) \), we know that
\[ \sqrt{n}(\hat{\theta} - \theta) \sim N(\overrightarrow{0}, \Sigma) \]

Denoting the fraction of individuals with observable characteristic \( w \) observed in frame \( d_j \)
by \( \alpha_{jw} = \frac{n_{jw}}{n} \), we can write the variance matrix as:

\[
\Sigma(\theta) = \\
\begin{pmatrix}
\frac{y_{0w_1}(1 - y_{0w_1})}{\alpha_{0w_1}} & \cdots & \frac{y_{0w_J}(1 - y_{0w_J})}{\alpha_{0w_J}} \\
\frac{y_{1w_1}(1 - y_{1w_1})}{\alpha_{1w_1}} & \cdots & \frac{y_{1w_J}(1 - y_{1w_J})}{\alpha_{1w_J}} \\
& \cdots & \\
\frac{p_{w_1}(1 - p_{w_1})}{\alpha_{w_1w_1}} & -p_{w_1}p_{w_2} & \cdots & -p_{w_1}p_{w_J} \\
-p_{w_1}p_{w_2} & p_{w_2}(1 - p_{w_2}) & \cdots & -p_{w_2}p_{w_J} \\
& \cdots & \cdots & \\
-p_{w_1}p_{w_J} & -p_{w_J}p_{w_1} & \cdots & p_{w_J}(1 - p_{w_J})
\end{pmatrix}
\]
where all blank entries of the $\Sigma$ matrix are zeroes.$^{127}$

The matching approach employs many different combinations of these primitive parameters. We begin with the weights for the subset of inconsistent choosers,

$$s_w \equiv p(w_i = w | c_i = 0) = \frac{Y_{1w} - Y_{0w}}{\sum_{w'=v}(Y_{1v} - Y_{ov})p_v}p_w$$

Carefully taking derivatives of this function and simplifying using the definition of $s_w$, we obtain the following for any $w$ and $w'$:

$$\frac{\partial s_w}{\partial Y_{0w'}} = -\frac{1\{w = w'\} - s_w}{\sum_{w'=v}(Y_{1v} - Y_{ov})p_v}p_w$$
$$\frac{\partial s_w}{\partial Y_{1w'}} = \frac{1\{w = w'\} - s_w}{\sum_{w'=v}(Y_{1v} - Y_{ov})p_v}p_w$$
$$\frac{\partial s_w}{\partial p_{w'}} = \frac{1\{w = w'\} - s_w}{p_w}$$

were $1\{\}$ is an indicator function equal to 1 when the expression inside the square brackets is true and zero otherwise. These three expressions can be used to generate the entire gradient of $s_w$.

Next we consider the weights for the subset of consistent choosers, $q_w \equiv p(w_i = w | c_i = 1) = \frac{p_{xw} + 1 - p_{yw}}{p_x + 1 - p_y}p_w$. Proceeding similarly to before, rewrite $R_w$ as

$$q_w = \frac{Y_{0w} + 1 - Y_{1w}}{\sum_{w'=v}(Y_{0v} + 1 - Y_{1v})p_v}p_w$$

Taking derivatives and simplifying, we obtain the following for any $w$ and $w'$:

$$\frac{\partial q_w}{\partial Y_{0w'}} = \frac{1\{w = w'\} - R_w}{\sum_{w'=v}(Y_{0v} + 1 - Y_{1v})p_v}p_w$$
$$\frac{\partial q_w}{\partial Y_{1w'}} = \frac{1\{w = w'\} - R_w}{\sum_{w'=v}(Y_{0v} + 1 - Y_{1v})p_v}p_w$$
$$\frac{\partial q_w}{\partial p_{w'}} = \frac{1\{w = w'\} - R_w}{p_w}$$

$^{127}$To be specific, we know that the off-diagonal elements in the first $2 \times J$ rows and the first $2 \times J$ columns, which govern the covariance of the various $\hat{y}_{jw}$ estimates, are zero because the estimation sample for every $\hat{y}_{jw}$ is distinct. We know that the entries of $\Sigma$ governing the covariance of $\hat{y}_{jw}$ and $p_w$, for some $j$ and $w$, are zero because of the unconfoundedness assumption.
Next we consider the estimators for the preferences of various subgroups. First, define the preferences of the inconsistent subgroup

\[ Y_N = \sum_w s_w Y_{cw} \]

Taking derivatives of this – which may be done more easily using several expressions derived above – we obtain the following for any \( w \):

\[
\frac{\partial Y_N}{\partial Y_{0w}} = s_w \frac{1 - Y_{cw}}{\bar{c}_w} + \frac{p_w}{1 - \bar{c}} (Y_N - Y_{cw})
\]

\[
\frac{\partial Y_N}{\partial Y_{1w}} = s_w \frac{Y_{cw}}{\bar{c}_w} - \frac{p_w}{1 - \bar{c}} (Y_N - Y_{cw})
\]

\[
\frac{\partial Y_N}{\partial p_w} = \frac{s_w}{p_w (1 - \bar{c})} [Y_{cw} - (1 - \bar{c}) Y_N]
\]

where \( \bar{c}_w \equiv Y_{0w} + 1 - Y_{1w} = E[c_i|w] \) and \( \bar{c} = \sum_w (Y_{0v} + 1 - Y_{1v}) p_v \). From these three expressions we construct the gradient of \( Y_N \).

Proceeding similarly for the full population, \( Y_{FP} = \sum_w p_w Y_{cw} \), we obtain, for any \( w \)

\[
\frac{\partial Y_{FP}}{\partial Y_{0w}} = p_w \frac{1 - Y_{cw}}{\bar{c}_w}
\]

\[
\frac{\partial Y_{FP}}{\partial Y_{1w}} = p_w \frac{Y_{cw}}{\bar{c}_w}
\]

\[
\frac{\partial Y_{FP}}{\partial p_w} = Y_{cw}
\]

which allows us to construct the gradient of \( Y_{FP} \).

We can also obtain the gradient of \( Y_c = \sum_w q_w Y_{cw} \) in terms of the primitive parameters of this model\(^{128}\). Taking derivatives of the expression for \( Y_c \) yields

\[
\frac{\partial Y_c}{\partial Y_{0w}} = q_w \frac{1 - Y_{cw}}{\bar{c}_w} + \frac{p_w}{\bar{c}} (Y_{cw} - Y_c)
\]

\(^{128}\)This part is not necessary to obtain a standard error on \( Y_c \), because we know how to obtain a simpler formula for the asymptotic variance of our estimator of \( Y_c \) using the result in the previous section of this Appendix. Doing it the hard way here yields an identical standard error estimate. The usefulness of the expressions derived here is that these expressions may be used to estimate the (asymptotic) covariance of, say, the estimators for \( Y_c \) and \( Y_N \), which is necessary for the statistical test of the null hypothesis of decision-quality independence against the alternative hypothesis of conditional decision-quality independence.
\[ \frac{\partial Y_c}{\partial Y_{1w}} = q_w \frac{Y_{cw}}{\bar{e}} - \frac{p_w}{\bar{e}} (Y_{cw} - Y_c) \]
\[ \frac{\partial Y_c}{\partial p_w} = \frac{q_w}{p_w \bar{e}} (Y_{cw} - \bar{Y}_c) \]

Using all of the above expressions, we can generate a gradient of each parameter of the matching-on-observables models. Putting all these expressions together, we construct a gradient matrix of the form:

\[ G(\theta) = (\nabla s_{w1}, \ldots, \nabla s_{wj}, \nabla q_{w1}, \ldots, \nabla q_{wj}, \nabla Y_N, \nabla Y_{FP}, \nabla Y_c) \]

To be clear, each of the columns of \( G(\theta) \) is the gradient of a particular (nonlinear) function of the primitive parameters of the model.

We can now estimate the full variance-covariance matrix of all the parameters \( (s_{w1}, \ldots, s_{wj}, q_{w1}, \ldots, q_{wj}, Y_N, Y_{FP}, Y_c) \) by

\[ \hat{V}(\hat{\theta}) = \frac{1}{n} G^\prime \Sigma G \]

where \( G \) and \( \Sigma \) are evaluated at \( \hat{\theta} \). The square root of the diagonals of the matrix \( \hat{V}(\hat{\theta}) \) will be asymptotically correct standard errors for the parameter estimates themselves. The off-diagonal elements are the estimated covariance of different estimates, which are useful for tests of hypotheses involving more than one parameter of the model, such as tests of decision-quality independence in this framework.

**Standard Errors for Proposition 4**

Using similar notation to before, let \( Y_{jk} \equiv E[y_{ijk}] \) for \( j = 0, 1 \) and \( k = h, l \), and denote the estimator for each population moment by \( \bar{y}_{jk} \). Similarly to the previous section, we begin by noting that
\[
\sqrt{n} \left( \begin{pmatrix} \bar{y}_{00} \\ \bar{y}_{0l} \\ \bar{y}_{10} \\ \bar{y}_{1l} \end{pmatrix} - \begin{pmatrix} Y_{00} \\ Y_{0l} \\ Y_{10} \\ Y_{1l} \end{pmatrix} \right) \overset{d}{\sim} \mathcal{N}(0, \Sigma)
\]

where \( \Sigma \) is a diagonal matrix with entries of the form \( \frac{1}{\alpha_{jk}} Y_{jk} (1 - Y_{jk}) \).

The new statistic in Proposition 4 is

\[
Y_s = \frac{Y_{0h} - Y_{0l}}{Y_{1l} - Y_{0l} - (Y_{1h} - Y_{0h})}
\]

which we can estimate consistently with:

\[
\hat{Y}_s = \frac{\bar{y}_{0h} - \bar{y}_{0l}}{\bar{y}_{1l} - \bar{y}_{0l} - (\bar{y}_{1h} - \bar{y}_{0h})}
\]

Using the delta method, we obtain \( V(\hat{Y}_s) \approx \frac{1}{n} \nabla Y_S' \Sigma \nabla Y_s \), where \( \nabla Y_S \) is evaluated at \((Y_{0h}, Y_{0l}, Y_{1h}, Y_{1l})\). Taking the gradient of (3.58) gives

\[
\nabla Y_s = \begin{pmatrix} \frac{Y_{1l} - Y_{1h}}{(\Delta c)^2}, \frac{Y_{1h} - Y_{1l}}{(\Delta c)^2}, \frac{Y_{0h} - Y_{0l}}{(\Delta c)^2}, \frac{Y_{0l} - Y_{0h}}{(\Delta c)^2} \end{pmatrix}'
\]

where \( \Delta c = Y_{1l} - Y_{0l} - (Y_{1h} - Y_{0h}) \). Plugging this into the formula for \( V(\hat{Y}_s) \) and simplifying yields

\[
V(\hat{Y}_s) = \left( \frac{1 - Y_s}{\Delta c} \right)^2 [V(\bar{y}_{0h}) + V(\bar{y}_{0l}) + \left( \frac{Y_s}{\Delta c} \right)^2 [V(\bar{y}_{1h}) + V(\bar{y}_{1l})]]
\]

where \( V(\bar{y}_{jk}) = \frac{1}{n_{jk}} Y_{jk} (1 - Y_{jk}) \). Replacing each \( Y_{jk} \) with the estimator \( \bar{y}_{jk} \), we obtain a consistent estimate of the asymptotic variance of \( \hat{Y}_s \). Note that when \( \Delta c \) is small, the variance of this estimator can be quite large, reflecting a familiar facet of instrumental variables estimation.
Bibliography


171


