Identification and Estimation of Labor Supply Elasticities from Kinked Budget Sets *

Deniz Dutz Morten Håvarstein

Magne Mogstad Alexander Torgovitsky

April 1, 2026

Abstract

We study the identification of labor supply elasticities from kinked budget sets in a model with income effects and individual heterogeneity in the elasticities. We provide point and partial identification results for compensated elasticities, uncompensated elasticities, and income effects. We use administrative data to apply our results to the Norwegian tax system, which exhibits a kink for the self-employed. There is clear bunching around the kink point, suggesting that the self-employed respond to the change in incentives created by the kink. We find that the bounds are often tight even under weak assumptions. Our results show that uncompensated elasticities are close to zero and compensated elasticities are sufficiently small to conclude that the excess burden of taxation is low.

Keywords: bunching, income effects, kinked budget sets, labor supply elasticities, partial identification.

JEL classification: C14, C18, H24, J22

*Dutz and Torgovitsky: Kenneth C. Griffin Department of Economics, University of Chicago. Håvarstein: Kenneth C. Griffin Department of Economics, University of Chicago; Department of Economics, University of Oslo. Mogstad: Kenneth C. Griffin Department of Economics, University of Chicago; Statistics Norway; NBER. Håvarstein and Mogstad acknowledge financial support from Norwegian Fiscal Studies and the Norwegian Research Council. Torgovitsky acknowledges support from National Science Foundation grant SES-1846832.

1 Introduction

Discontinuous incentives can lead economic agents to make choices that cluster or "bunch." This type of bunching behavior has been documented across a wide range of settings, including labor earnings (Saez, 2010), firm profits (Best et al., 2015), intertemporal savings decisions (Best et al., 2019), and inheritance (Glogowsky, 2021); see Kleven (2016) and Bertanha et al. (2024) for surveys.

Saez (2010) showed how to use bunching in earnings around kinks in marginal tax rates to estimate labor supply elasticities. His approach has been applied extensively in the empirical literature; see, for example, Chetty et al. (2011), Bastani and Selin (2014), and Mortenson and Whitten (2020). Saez's estimation approach assumes that there are no income effects and that compensated elasticities are homogeneous across individuals.

We provide new results on identification and estimation when the assumptions of no income effects and/or homogenous elasticities are removed. The model of labor supply that we consider nests models commonly used in public finance, but allows for both compensated elasticities and income effects to be random variables that vary across the population. The model generates bunching behavior when a progressive increase in marginal tax rates creates a kink in the budget set. The way in which the distribution of earnings is affected by the introduction of this kink depends on the unknown distribution of labor supply elasticities.

Our main results are bounds on average labor supply elasticities that are sharp, meaning they extract *all* of the information in the data and model. We assume that the researcher observes the distribution of earnings under the kinked system for all individuals at or above the kink and either observes or estimates some portion of the distribution of earnings in a system without the kink. The bounds that we derive utilize these two distributions together with the economic structure of the labor supply model. The bounds are analytic, making estimation and inference straightforward. We also allow the researcher to impose prior restrictions that rule out extreme elasticities. As part of our analysis, we provide sharp testable implications for the model under various sets of assumptions.

We apply our results to Norwegian administrative data on the earnings of the self-employed, where we find clear evidence of bunching around the kink in the top earnings bracket. We test and reject the assumptions of no income effects and homogeneous compensated elasticities. A model with heterogeneous compensated elasticities and no income effects is not rejected and produces remarkably tight bounds. Allowing for heterogeneous income effects widens these bounds modestly. Under conservative assumptions, we estimate that average compensated elasticities are at most 0.161 allowing for large income effects, and at most 0.073 with no income effects. We also find evidence that uncompensated elasticities are relatively close to zero. A direct implication of our findings is that the excess burden of taxation is small.

In Section 2, we introduce the labor supply model and the tax system. We initially focus our attention on kinks in the top bracket, which is an easier case to analyze and aligns with our empirical application. Later in the paper, we generalize our results to interior brackets.

We begin our identification analysis in Section 3 by revisiting Saez's (2010) estimation approach, which assumed no income effects and a homogeneous compensated elasticity. Saez only uses the "bunching quantile" of the pre-kink distribution, which under his assumptions corresponds to the pre-kink earnings of the marginal buncher. Because there are no income effects, the compensated elasticity of this marginal buncher—*the* compensated elasticity under homogeneity—can be inferred from their earnings response. We show that Saez's approach continues to work if income effects are non-zero but known and homogeneous. We also show that using more of the pre-kink distribution allows one to point identify both a homogeneous compensated elasticity and a homogeneous income effect. The homogeneous model is in fact overidentified, an implication we use to construct a specification test, which rejects in our application.

We then consider models with heterogeneous compensated elasticities and/or income effects. The relevant target parameters in these models are generally only partially identified. In Section 4, we introduce the relevant machinery for studying partially identified models.

We use this machinery in Section 5, where we allow for heterogeneous compensated elasticities while continuing to assume that income effects are known and homogeneous. Having a heterogeneous compensated elasticity breaks Saez's approach because the bunching quantile no longer needs to reflect the behavior of the marginal buncher. We show that the identification problem can be recast as a contaminated data problem, in which the observed pre-kink earnings distribution is a mixture of the pre-kink conditional earnings distributions of bunchers and non-bunchers. These conditional distributions are not directly observed, because they depend on post-kink bunching behavior. However, the shares of bunchers and non-bunchers are observed. This allows us to apply a trimming argument originally developed by Horowitz and Manski (1995), which we show how to repurpose to bound average compensated elasticities for bunchers, non-bunchers, and the combined group.

In Section 6, we allow for both compensated elasticities and income effects to be heterogeneous. This change complicates the trimming argument because it breaks the direct link between earnings responses and compensated elasticities. However, there is still an indirect link through Engel aggregation, which implies that income effects cannot be arbitrarily large. We show how to use this logic to derive sharp bounds on average income effects, average compensated elasticities, and average uncompensated elasticities, both by bunching status and for the combined group.

In Section 7, we generalize our arguments to apply to interior tax brackets. This complicates the analysis because it raises the possibility of bracket switching, a possibility which does not arise when analyzing a top bracket under standard preferences. We show how to solve this problem by focusing attention on an identifiable subpopulation that will not switch brackets. Our analysis also provides a way to study top earnings brackets when using less of the pre-kink distribution, which can be useful to allay concerns about extrapolation. In Section 8, we use these results in our empirical application, which we find is quite robust to extrapolation.

We provide a brief conclusion in Section 9 that summarizes our central findings. Proofs for all results can be found in the appendix and supplemental appendix.

Our paper contributes to a methodological literature about the use of kinks and notches dating back to at least Burtless and Hausman (1978). See Bertanha et al. (2024) for a recent survey. Some of this literature has focused on learning about compensated elasticities when the pre-kink earnings distribution is unknown (Blomquist et al., 2021), while others also consider the problem of extrapolating from one observed tax system to another unobserved one (Bertanha et al., 2022, 2023). Our focus is on what can be determined about labor supply elasticities *given* knowledge of the earnings distributions when there are income effects and elasticities are heterogeneous across observationally-identical agents. Pollinger (2025) starts from the same premise, but considers identification of a model with an extensive margin and homogeneous elasticities. Goff (2024) considers identification of the earnings response (treatment effects) from a tax change around the kink, whereas our focus is on identifying elasticity parameters that map into key economic quantities, such as the excess burden of taxation.

2 Model and data

In this section, we lay out the model of labor supply that we use throughout the analysis. We specify the tax system and show how it leads to bunching. We define the identification problem when using different features of the earnings distribution. We also introduce the Norwegian income data to which we apply our results.

2.1 Labor supply

Consider a model of possibly-heterogeneous individuals choosing consumption C and labor supply Y. Following Feldstein (1999) and Saez (2010), we take Y to be pre-tax earnings to capture labor supply choices along multiple margins of adjustment. Individuals have utility $U(C,Y)$ over consumption and labor supply. We assume that U is twice continuously differentiable, strictly increasing in C, strictly decreasing in Y, has a strictly negative mixed partial derivative, and is strictly concave as a function of both arguments. These assumptions imply that preferences are convex.

We assume that individuals face a progressive tax system that is piecewise linear. We initially focus on individuals with earnings in the top bracket, defined by a threshold $\bar{Y}$, but we extend our arguments to interior brackets in Section 7. As shown by Hall (1973), convex preferences ensure that individuals who choose earnings Y above this threshold $\bar{Y}$ behave as if they were facing an appropriately-defined "virtual" linear tax system, even if the actual tax system is nonlinear. Letting the marginal tax rate of this virtual system be denoted by t and the virtual transfer by R, individuals maximize utility subject to the budget constraint imposed by the tax system:

$$\max_{C,Y} U(C,Y) \quad \text{subject to} \quad C = (1-t)Y + R.$$

The assumptions on U ensure that the individual's problem has a smooth, unique solution, which we denote as $Y^\star(1-t,R)$.

We denote the uncompensated (Marshallian) elasticity of labor supply with respect to the take-home rate as

$$\epsilon^u(1-t,R) \equiv \frac{1-t}{Y^*(1-t,R)} \frac{\partial Y^\star(1-t,R)}{\partial(1-t)}.$$

The income effect is defined as

$$\eta(1-t,R) \equiv (1-t)\frac{\partial Y^\star(1-t,R)}{\partial R}.$$

Slutsky's equation implies that the compensated (Hicksian) elasticity of labor supply with respect to the take-home rate is then $\epsilon(1-t,R) \equiv \epsilon^u(1-t,R) - \eta(1-t,R) \geq 0$.[1] Our assumptions on utility imply that consumption and leisure are normal goods, leading to the Engel aggregation condition (adding-up constraint) that $\eta(1-t,R) \in [-1,0]$.

[1] The compensated earnings function depends on the marginal tax rate and the level of utility. Letting $U^\star(1-t,R)$ denote the indirect utility function, $\epsilon(1-t,R)$ is the elasticity of the compensated earnings function with respect to the net-of-tax rate, evaluated at $1-t$ and $U^\star(1-t,R)$.

We assume the uncompensated (Marshallian) earnings function under a linear virtual tax system satisfies

$$\log Y^{\star}(1-t, R) = \log\beta_0 + \beta_t \log(1-t) + \beta_R \phi(R), \tag{EF}$$

for some strictly increasing, twice differentiable, and (weakly) concave function ϕ. A special case of (EF) with $\beta_R = 0$ follows from the assumption that

$$U(C, Y) = C - \frac{\beta_0}{1 + 1/\beta_t}\left(\frac{Y}{\beta_0}\right)^{1+1/\beta_t},$$

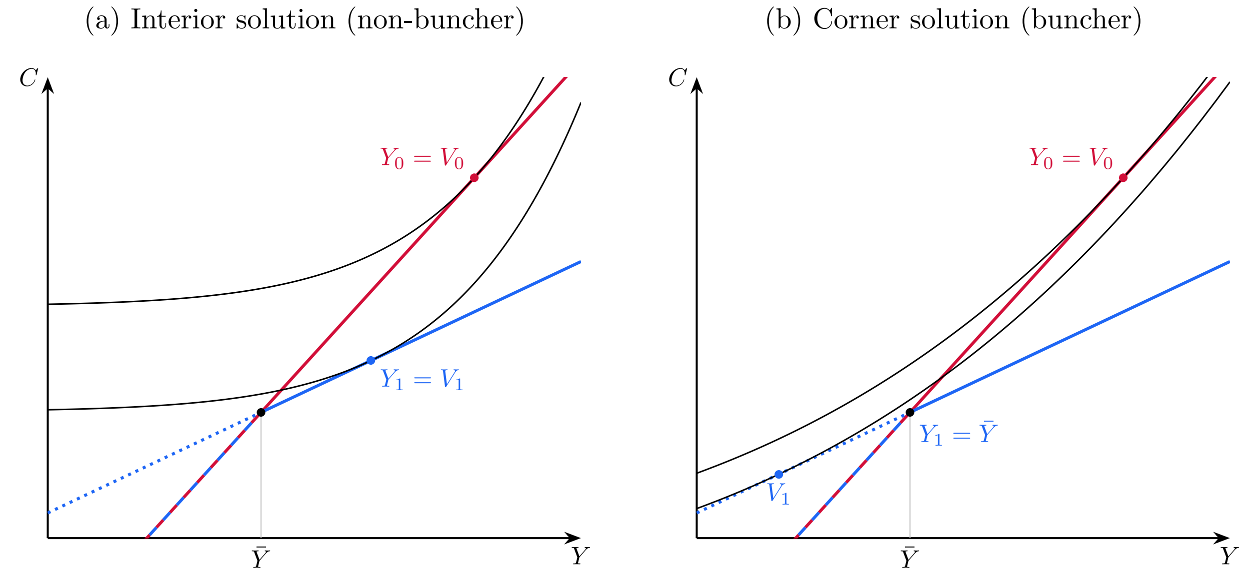
which is the specification of utility used by Saez (2010) and many other authors. More generally, earnings functions like (EF) are common in the literature; see, for example, Auten and Carroll (1999), Gruber and Saez (2002), Saez et al. (2012), or Kleven and Schultz (2014). All of these authors estimate earnings functions where log earnings are linear in $\log(1-t)$, although they vary in their choice of ϕ. For example, Kleven and Schultz (2014) use $\phi(R) = \log(R)$, so that log earnings are linear in $\log R$. Our theoretical results apply for any ϕ that satisfies the shape constraints. In our application, we set $\phi(R) = R$ so that it is well-defined for $R = 0$.

2.2 The effect of kinks in the tax system

Figure 1 depicts a piecewise linear tax system with marginal tax rate t_0 for earnings below the threshold ($\bar{Y}$) and marginal tax rate $t_1 > t_0$ for earnings above the threshold. We call this tax system 1, which we think of as the status quo system with a kink at the threshold. Figure 1 also shows an alternative linear tax system that has marginal tax rate t_0 throughout. We call this tax system 0, which we think of as a "pre-kink" system. Hall's virtual linear tax system under system 1 is shown by the dashed blue line in Figure 1. If R_0 denotes the virtual transfer under system 0, then the virtual budget set has virtual transfer $R_1 \equiv R_0 + \bar{Y}(t_1 - t_0)$.

We compare individual behavior in the two systems by considering a hypothetical change from system 0 to system 1. This change would not affect the choices of individuals with earnings below the threshold, because their original bundle under system 0 would still be available under system 1, while bundles above the kink have become less attractive. We only consider individuals above the kink under system 0. These individuals will adjust their earnings in response to the increased marginal tax rate, but they will remain above the kink, because preferences are convex. Following Hall's (1973) insight, their choices can be determined by considering their behavior under the virtual *linear* budget set for system 1, which is shown in blue in Figure 1. The next

Figure 1: Illustration of bunchers and non-bunchers



Notes: The system 0 *budget set,* $C = (1-t_0)Y + R_0$*, is shown in red and the system* 1 *budget set is shown in blue. The budget sets overlap to the left of* $\bar{Y}$*, where they have slope* $1-t_0$*. The slope of the system* 1 *budget set to the right of* $\bar{Y}$ *is* $1-t_1 < 1-t_0$*. The dashed blue line to the left of* $\bar{Y}$ *denotes the virtual linear budget set implied by system* 1*, which has budget constraint* $C = (1-t_1)Y + R_1$ *with* $R_1 \equiv R_0 + \bar{Y}(t_1 - t_0)$*. The thin black lines are indifference curves. The non-buncher chooses an interior solution of* $V_1 = Y_1 \geq \bar{Y}$*, while the buncher has* $V_1 < \bar{Y}$*, so chooses the corner solution* $Y_1 = \bar{Y}$*. System 0 is already linear, so* $V_0 = Y_0$*.*

proposition describes how their choices under the virtual linear budget set for system 1 would be related to their choices under system 0.

Proposition 1. Consider an individual who would choose $V_0 \equiv Y^\star(1-t_0, R_0)$ under the linear tax system 0. Let $V_1 \equiv Y^\star(1-t_1, R_1)$ denote their earnings choice under a hypothetical virtual linear system with constant marginal tax rate $t_1 > t_0$ and transfer $R_1 \equiv R_0 + \bar{Y}(t_1 - t_0)$. Let $v_d \equiv \log V_d$. Then

$$v_1 = v_0 - \epsilon\tau - \eta\pi(v_0) \equiv \nu(v_0, \epsilon, \eta), \tag{VE}$$

where ϵ and η are elasticities evaluated at $(1-t_0, R_0)$, $\tau \equiv \log(1-t_0) - \log(1-t_1)$ is the change in log take-home rates, and

$$\pi(v_0) \equiv \tau - \frac{\phi(R_1) - \phi(R_0)}{(1-t_0)\exp(v_0)\phi'(R_0)}$$

is a strictly increasing function of v_0 with $\pi(v_0) \in (0, \tau)$ for all $v_0 \geq \bar{y} \equiv \log(\bar{Y})$.

The virtual earnings choices under system 1 given in (VE) reflect a combination of both income (η) and substitution effects (ϵ), and depend on earnings choices v_0 under system 0. Note that because system 0 is itself linear, the virtual and actual earnings choices are the same: $v_0 = y_0$.[2] In contrast, virtual earnings choices under system 1 are only feasible if $v_1 \geq \bar{y}$: those with $v_1 < \bar{y}$ cannot choose $y_1 = v_1$ under the actual tax system. Because preferences are convex, their best feasible option is to "bunch" at the kink by choosing $y_1 = \bar{y}$.[3] Log earnings under system 1 are therefore

$$y_1 = \begin{cases} v_1, & \text{if } v_1 > \bar{y} \quad \text{(interior solution — non-bunchers)} \\ \bar{y}, & \text{if } v_1 \leq \bar{y} \quad \text{(corner solution — bunchers).} \end{cases} \tag{BU}$$

The combination of (VE) and (BU) is the foundation of our empirical analysis. Figure 1a illustrates the behavior of a non-buncher, while Figure 1b illustrates the behavior of a buncher.

2.3 The empirical problem

The researcher observes the distribution of y_1, denoted G_1. They also observe or estimate some portion of the distribution of y_0, denoted G_0. Their goal is to use this information to identify features of the distribution of the elasticity parameters, ϵ and η. The nature of the problem depends both on what features of G_0 are assumed to be known (or knowable), and on whether the elasticity parameters are allowed to be stochastic (heterogeneous) or are assumed to be constant (homogeneous).

As discussed in the previous section, the only individuals whose choices are affected by the reform are those with $y_0 > \bar{y}$ and all individuals who choose $y_1 \geq \bar{y}$ also have $y_0 > \bar{y}$. This implies that the only empirically relevant portion of the distribution of earnings is where $y_0 > \bar{y}$, or, equivalently, $y_1 \geq \bar{y}$. We condition on this event implicitly in the notation, so that G_0 and G_1 are the distributions of y_0 and y_1 conditional on the event $y_0 > \bar{y}$ or the equivalent event $y_1 \geq \bar{y}$. Likewise, all expectation and probability operators should be interpreted as implicitly conditional on this event.[4] We assume throughout our analysis that G_0 is strictly increasing and absolutely continuous, so that it admits a density. We make the same assumption on the distribution of y_1 implied by G_1, conditional on $y_1 > \bar{y}$, noting that G_1 has a mass point at $\bar{y}$ due to bunching.[5]

[2] The distinction between v_0 and y_0 becomes important when considering interior brackets in Section 7.

[3] Suppose preferences are strictly convex and an individual with $v_1 < \bar{y}$ chooses $y_1 > \bar{y}$. Then there exists a feasible point in $[\bar{y}, y_1)$ that is strictly preferred to y_1, violating rationality.

[4] These notational distinctions become more important when analyzing interior brackets in Section 7.

[5] Assuming continuity makes many of our bounding results cleaner, but our arguments can be modified for discrete distributions as well.

We denote the size of this mass point—the proportion of bunchers—by $\bar{p} \equiv G_1(\bar{y})$. The bunching method developed by Saez (2010) uses the distribution of G_0 up to the $\bar{p}$th quantile, which we denote as $q_0(\bar{p})$ and refer to as the "bunching quantile." No other features of G_0 are used.

As we show ahead, there is a great deal of additional information contained in the rest of G_0. We generalize the bunching method to the case in which the researcher uses information in the distribution of G_0 up to $\bar{\bar{y}}_0$. Let $\bar{\bar{p}}_0 \equiv G_0(\bar{\bar{y}}_0)$ be the quantile associated with $\bar{\bar{y}}_0$. If $\bar{\bar{p}}_0 = 1$, the researcher is using the entirety of the distribution of G_0. The bunching method is nested as $\bar{\bar{p}}_0 = \bar{p}$. We also consider intermediate data methods that choose $\bar{\bar{y}}_0$ so that $\bar{\bar{p}}_0$ is somewhere between $\bar{p}$ and 1. As we show, setting $\bar{\bar{y}}_0$ to be larger—using more of the distribution of G_0—allows one to learn more about the underlying elasticities.

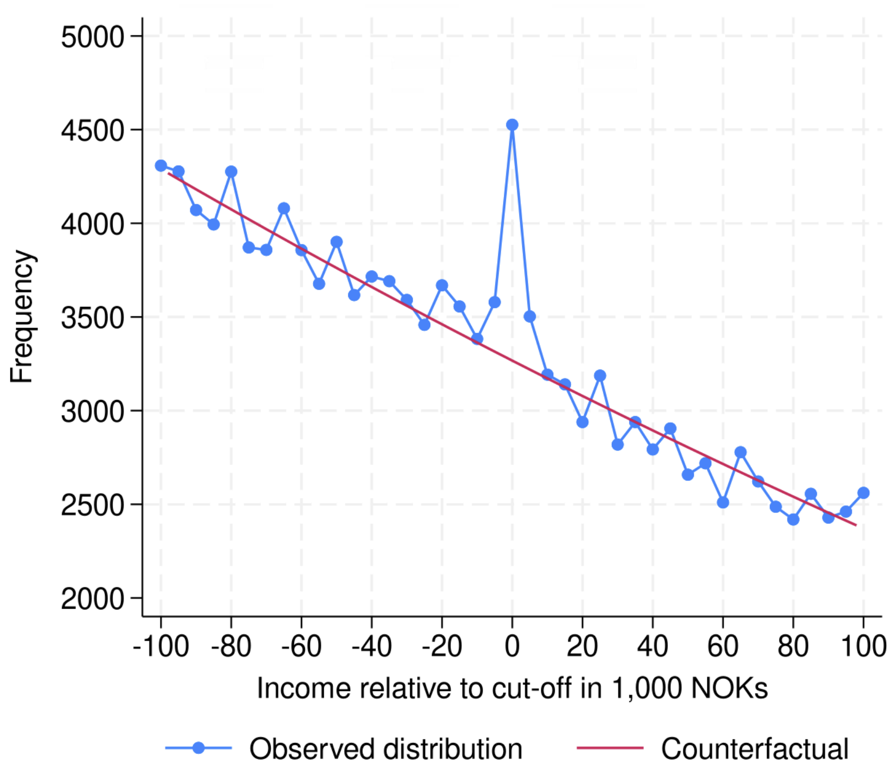
All of the methods we discuss require knowing the quantiles of G_0 up to $\bar{\bar{p}}_0$. Saez (2010) and Chetty et al. (2011) use parametric assumptions to estimate the bunching quantile $q_0(\bar{p})$ by extrapolating data from system 1 into a region around the kink point. Blomquist et al. (2021) show that these parametric assumptions are important because the counterfactual distribution is not non-parametrically identified without additional information. Bertanha et al. (2023) propose maximum likelihood-based estimators to jointly estimate G_0 and ϵ under no income effects and homogeneous ϵ. If G_0 is estimated by parametric extrapolation, then estimates of its tails might be unreliable. This sets up a trade-off that is controlled by $\bar{\bar{p}}_0$: a larger $\bar{\bar{p}}_0$ allows one to learn more, but risks unreliable extrapolation, while a smaller $\bar{\bar{p}}_0$ achieves the opposite.

Knowledge of G_0 could also come from sources that do not require parametric extrapolation. For example, it could come from an experimental evaluation such as the Negative Income Tax (e.g. Ashenfelter and Plant, 1990) or Jobs First (e.g. Bitler et al., 2006) experiments. It could also come from quasi-experimental methods. For example, Coles et al. (2022) use firms that face different tax kinks, Hungerman and Ottoni-Wilhelm (2021) use the distribution in states that have different tax laws, and Gelber et al. (2021) use the distribution of earnings for differently-aged workers who do not face a kink. Our analysis ahead is premised on knowing G_0 up to the $\bar{\bar{p}}_0$ quantile, but takes no stance on where this knowledge comes from, whether parametric extrapolation, an experimental or quasi-experimental evaluation, or some other method.

2.4 Data

We apply our results using Norwegian administrative data from 2006–2018. The data contain separate measures for wage, business, and capital income, as well as taxable

Figure 2: Actual and counterfactual earnings distribution around the kink



Notes: The figure plots the actual and counterfactual earnings distribution around the kink. Observations within 30,000 *NOK of the kink are excluded when estimating the counterfactual distribution. The counterfactual distribution is recovered using a similar approach as Chetty et al. (2011). See Appendix A for details.*

income. We focus our attention on prime-aged workers between the ages of 26 and 61 who are self-employed, defined as having business income greater than wage income in a given year. This restriction is motivated by earlier research that finds more bunching among the self-employed than among wage earners (Saez, 2010; Chetty et al., 2011; Bastani and Selin, 2014).[6] Appendix Table C.1 shows that the self-employed tend to earn about 10% more than the population, are about three years older on average, more likely to be male, and less likely to have a college degree.

Our focus is on the top two tax brackets. Appendix Figure C.2 shows that the income level for the top bracket as well as the marginal tax rates in the top two brackets

[6]One possible reason for this finding is that they can more precisely control their earnings. Another possibility is that they are better informed about the tax system. We refer to Chetty (2012), Kleven and Waseem (2013), and Kostøl and Myhre (2021) for discussions about how optimization and information frictions could affect bunching at kinks.

remained stable over the horizon we study. Like Saez (2010) and Chetty et al. (2011), we use this stability to gain precision by pooling the data across years and defining marginal tax rates above and below this average boundary by their average across years. The resulting kink is located at roughly 912,500 Norwegian Kroner (NOK), which was approximately $110,000 in 2018. The marginal tax rate changes at the kink from 0.436 to 0.466. Accounting for Norway's 25 percent value-added tax, we interpret this as $t_0 = 0.549$ and $t_1 = 0.573$.[7]

Figure 2 plots the earnings distribution in blue, which corresponds to the system 1 distribution G_1. A spike in the density of earnings gives clear visual evidence of bunching. We estimate the counterfactual system 0 earnings distribution, G_0, by closely following the extrapolation method used by Chetty et al. (2011). Their method assumes that the density of G_0 is proportional to the actual distribution above the kink and that it follows a 7th-order polynomial within 30,000 NOK of the kink. We formalize their assumptions and estimators in Appendix A. The resulting estimate of G_0 is shown by the red line in Figure 2. The counterfactual density is smooth and closely tracks the earnings distribution below the kink, suggesting that the extrapolation is plausible.

Chetty et al. (2011) measure the size of the bunch by calculating the "excess mass," defined as the share of bunchers $\bar{p}$ relative to the share of individuals with Y_0 close to $\bar{Y}$. If we define close as 5,000 NOK, then their definition of excess mass is $\bar{p}/(G_0(\bar{Y} + 5000) - G_0(\bar{Y}))$, which is approximately 0.79 in our data. The standard error is 0.003, so we strongly reject the null hypothesis that there is no bunching at $\bar{Y}$.

3 Homogeneous substitution and income effects

In this section, we consider identification when ϵ and η are homogeneous (non-stochastic) parameters that do not vary across individuals. The primary target parameters are clear in this case: ϵ and η. From these parameters, we can also compute the implied uncompensated elasticity $\epsilon^u = \epsilon + \eta$.

3.1 Identification using the bunching quantile

With homogeneous elasticities, the virtual earnings equation (VE) defines a deterministic relationship between v_1 and v_0 via $v_1 = \nu(v_0, \epsilon, \eta)$. The function $\nu(v_0, \epsilon, \eta)$ is strictly increasing in v_0 because $\pi(v_0)$ is strictly increasing and $\eta \leq 0$. Together with

[7] We make the adjustment for the value-added tax (VAT) as follows. Let the nominal top-income tax rate be denoted t_y and VAT by t_c. The worker's budget constraint can be written as $(1 + t_c)C = (1 - t_y)Y + R$, which means one dollar of pre-tax earnings converts to $(1-t)$ dollars of consumption, with $t = 1-(1-t_y)/(1+t_c)$. Substituting for $t_y = 0.436$ or $t_y = 0.466$ and $t_c = 0.25$ produces $t_0 = 1 - (1 - 0.436)/(1 + 0.25) \approx 0.549$ and $t_1 = 1 - (1 - 0.466)/(1 + 0.25) \approx 0.573$.

the bunching behavior described in (BU), this means that there is a maximum earnings level $\nu^{-1}(\bar{y}, \epsilon, \eta)$ that an individual who chooses to bunch under system 1 could have under system 0. This implies that the proportion that bunch at the kink satisfies

$$\bar{p} = \mathbb{P}[\nu(v_0, \epsilon, \eta) \leq \bar{y}] = \mathbb{P}[v_0 \leq \nu^{-1}(\bar{y}, \epsilon, \eta)] = \mathbb{P}[y_0 \leq \nu^{-1}(\bar{y}, \epsilon, \eta)], \tag{1}$$

where the last equality follows because system 0 is itself linear, implying that $v_0 = y_0$. From (1), we conclude that the bunching quantile $q_0(\bar{p})$ satisfies

$$q_0(\bar{p}) = \nu^{-1}(\bar{y}, \epsilon, \eta) \quad \text{or} \quad \nu(q_0(\bar{p}), \epsilon, \eta) = \bar{y}. \tag{2}$$

Using (VE) then provides the following conclusion.

Proposition 2. Suppose that $\bar{\bar{p}}_0 = \bar{p}$. If η is assumed to be known, then ϵ is point identified:

$$\epsilon = \frac{q_0(\bar{p}) - \bar{y} - \eta\pi(q_0(\bar{p}))}{\tau}. \tag{3}$$

If η is unknown, then ϵ is not point identified, but it is bounded:

$$\frac{q_0(\bar{p}) - \bar{y}}{\tau} \leq \epsilon \leq \frac{q_0(\bar{p}) - \bar{y} + \pi(q_0(\bar{p}))}{\tau}. \tag{4}$$

These bounds are sharp given only knowledge of G_1 and $G_0(y)$ for $y \leq q_0(\bar{p})$.

Proposition 2 nests the case of no income effects ($\eta = 0$), so it both formalizes and generalizes the identification argument in Saez (2010). The argument shows that the marginal buncher has earnings $q_0(\bar{p})$ in system 0. The labor supply response for the marginal buncher is not censored, so their virtual response $v_1 - v_0 = \nu(y_0, \epsilon, 0) - y_0$ is the same as their observed response $\bar{y} - q_0(\bar{p})$. Using (VE) to substitute for the virtual earnings produces the expression for the compensated elasticity provided in (3), which is identified if η is assumed to be known. Assuming elasticities are homogeneous means that this compensated elasticity for the marginal buncher also applies to all other individuals.

The bounds in (4) under unknown η arise from considering all values of η between the range of 0 and -1 allowed for by the Engel aggregation condition. The sharpness statement in the second part of Proposition 2 implies that ϵ is not point identified from the bunching quantile without an assumption about income effects, η. Assuming away income effects ($\eta = 0$) is one possibility, but any other assumed value of η would also imply a value of ϵ. Considering the range of income effects consistent with Engel aggregation gives the bounds shown in (4). The lower bound is attained when $\eta = 0$, while the upper bound is attained when $\eta = -1$.

3.2 Estimates using the bunching quantile

The first two rows of Table 1 report point estimates and bounds on ϵ, ϵ^u, and η using only the bunching quantile. Under the assumption of no income effects ($\eta = 0$), the compensated elasticity is point identified and estimated to be 0.079. This estimate is similar to other bunching estimates for the self-employed in Nordic countries. Bastani and Selin (2014) find estimates ranging from 0.020 to 0.073 in Sweden. Chetty et al. (2011) find an estimate of 0.1 at an interior kink in the Danish tax system.

Allowing for income effects breaks point identification, but the compensated elasticity remains tightly bounded between 0.079 and 0.110. The reason the bounds are tight is that the *average* tax rate of the marginal buncher—an individual who must be close to the kink given the homogeneity assumptions—barely changes as the tax system changes from system 0 to system 1. We conclude that assuming no income effects has little impact on the estimated compensated elasticities when these elasticities are assumed to be homogeneous.

Table 1 also reports bounds on the uncompensated elasticity when using only the bunching quantile. The upper bound on ϵ^u of 0.079 is the lower bound on ϵ, which rules out large uncompensated elasticities. The lower bound on ϵ^u of -0.890 is not informative, but it is sharp. We conclude that it is more difficult to learn about uncompensated elasticities than it is to learn about compensated elasticities when using only the bunching quantile. The intuition is that the uncompensated earnings elasticity measures the earnings response to tax reforms that change the marginal and average tax rate by the same amount. However, the introduction of a kink barely changes the average tax rate for the marginal buncher, implying that the observed earnings response is much closer to a compensated earnings response than an uncompensated one.

3.3 Using more data

Proposition 2 showed that the bunching method with $\bar{\bar{p}}_0 = \bar{p}$ does not provide enough information to separately identify ϵ and η: there is a single equation (3) with two unknowns, ϵ and η. Setting $\bar{\bar{p}}_0 > \bar{p}$ allows for identification of both unknowns by providing equations for multiple quantiles $p \geq \bar{p}$. The underlying reasoning is similar:

$$p = \mathbb{P}[y_0 \leq q_0(p)] = \mathbb{P}[\nu(y_0, \epsilon, \eta) \leq \nu(q_0(p), \epsilon, \eta)] = \mathbb{P}[y_1 \leq \nu(q_0(p), \epsilon, \eta)], \qquad (5)$$

Table 1: Elasticities assuming homogeneity

$\bar{\bar{p}}_0$	Income effects	Compensated elasticity		Uncompensated elasticity		Income effect	
		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Bunching quantile ($\bar{p}$)	×	0.079	$(0.078, 0.079)$				
Bunching quantile ($\bar{p}$)	✓	$[0.079, 0.110]$	$(0.078, 0.110)$	$[-0.890, 0.079]$	$(-0.890, 0.078)$	$[-1.000, 0.000]$	
0.05	✓	0.080	$(0.079, 0.080)$	0.043	$(0.034, 0.051)$	-0.037	$(-0.046, -0.028)$
0.10	✓	0.080	$(0.079, 0.080)$	0.047	$(0.045, 0.048)$	-0.033	$(-0.035, -0.031)$
0.20	✓	0.079	$(0.078, 0.079)$	0.061	$(0.061, 0.062)$	-0.017	$(-0.018, -0.017)$
1.00	✓	0.082	$(0.082, 0.083)$	0.016	$(0.016, 0.017)$	-0.066	$(-0.066, -0.065)$

Notes: This table shows estimates of elasticities under the assumption that they are homogeneous. The first row assumes no income effects, while the remaining rows allow for income effects. Estimates and 95% confidence intervals are presented when using data up until different choices of $\bar{\bar{p}}_0$ ranging from the bunching quantile, to the 5th, 10th, and 20th quantiles of y_0, up until using full knowledge of the distribution of y_0 (100th quantile). Confidence intervals are constructed using the method of Imbens and Manski (2004) with the bootstrap to account for estimating the pre-kink distribution.

where the second equality uses the strict monotonicity of ν in $v_0 = y_0$ and the third holds because $q_0(p) \geq \bar{y}$ if $p \geq \bar{p}$. Let $q_1(p)$ denote the pth quantile of the earnings distribution under system 1. Then (5) together with (VE) implies

$$q_1(p) = \nu(q_0(p), \epsilon, \eta) \quad \Leftrightarrow \quad \tau\epsilon + \eta\pi(q_0(p)) = q_0(p) - q_1(p). \tag{6}$$

Knowledge of $q_0(p)$ and $q_1(p)$ at two different quantiles provides two equations that can be solved for ϵ and η.

Proposition 3. Suppose that $\bar{\bar{p}}_0 > \bar{p}$, so that $q_d(p'), q_d(p'')$ are known for at least two distinct $p', p'' \geq \bar{p}$ and both $d = 0, 1$. Then

$$\eta = \frac{(q_0(p'') - q_1(p'')) - (q_0(p') - q_1(p'))}{\pi(q_0(p'')) - \pi(q_0(p'))} \tag{7}$$

$$\text{and} \quad \epsilon = \frac{q_0(p') - q_1(p') - \eta\pi(q_0(p'))}{\tau}. \tag{8}$$

Proposition 3 shows that if ϵ and η are homogeneous, then they are separately identified if $\bar{\bar{p}}_0 > \bar{p}$. The intuition is that labor supply responses to system 1 depend on how far an individual's system 0 earnings are from the kink. If they are close to the kink, then their response is driven more by the compensated elasticity, while if they are far from the kink, their response is driven more by their uncompensated elasticity. The exact trade-off is governed by π, which measures the proportional distance from the kink in levels. If elasticities are homogeneous, then comparing responses for any two quantiles is enough to isolate both ϵ and η.

This reasoning also implies that ϵ and η are heavily overidentified: *any* two quantiles p', p'' should satisfy (7)–(8). Multiple quantiles can be combined in estimation. Given estimators $\hat{q}_d(p)$ for $d = 0, 1$, one can estimate ϵ and η by solving the least squares problem suggested by (6):

$$(\hat{\epsilon}, \hat{\eta}) = \arg\min_{\epsilon,\eta} \sum_{j=1}^{n} \left[(\hat{q}_0(p_j) - \hat{q}_1(p_j)) - \tau\epsilon - \eta\pi(\hat{q}_0(p_j)) \right]^2 \tag{9}$$

where $p_1, \ldots, p_n$ is a grid of quantiles below $\bar{\bar{p}}_0$. This is simple to implement as an ordinary least squares estimator.[8] It also provides a testable implication of the model, namely that the sum of squared residuals of (9) is zero. This implication is sharp in the sense that if some $\epsilon \geq 0$ and $\eta \in [-1, 0]$ satisfies (6) for all quantiles p, then (ϵ, η) is consistent with the model and maps the marginal distribution of y_0 to the marginal distribution of y_1.

The last four rows of Table 1 report point estimates of ϵ, η, and ϵ^u based on (9) using different values for $\bar{\bar{p}}_0$. The point estimates of the compensated elasticity remain close to 0.080 regardless of how much data is used. The estimated uncompensated elasticities vary between 0.016 and 0.061 for different values of $\bar{\bar{p}}_0$. An uncompensated elasticity close to zero is consistent with macroeconomic time-series evidence that hours worked do not increase as wages rise (see, for example, Kimball and Shapiro, 2008; Boppart and Krusell, 2020). The estimated income effects range from -0.066 to -0.017. These are smaller than those obtained from using lottery winnings by Cesarini et al. (2017) and Golosov et al. (2024). However, these studies estimate a combination of intensive and extensive margin responses, so their estimates are not directly comparable to ours.

In Appendix SA.1, we derive expressions for the revenue-maximizing tax rate and the excess burden of taxation as functions of the labor supply elasticities. Our compensated elasticity estimate with $\bar{\bar{p}}_0 = 0.1$ implies an excess burden of taxation of 11 percent. This means that the cost of increasing the marginal tax rate while keeping taxpayers equally well off is 11 cents per dollar raised. The revenue-maximizing tax rate on incomes above $\bar{Y}$ is 0.86, which is much higher than the status quo tax rate.

The model with homogeneous substitution and income effects is overidentified when using data above the bunching quantile. We implement a bootstrap test of the null hypothesis that the model is not misspecified when using different amounts of data.[9]

[8] Two modifications of this estimator are worth mentioning. First, we could add the constraints $\epsilon \geq 0$ and $\eta \in [-1, 0]$ to the problem. In our setting, these constraints are far from binding, so would make no difference. Second, instead of using equal weighting, the quantiles could be combined using an efficient minimum distance weighting. Working out the efficient weighting would be quite complicated for the choice of $\hat{q}_0(p)$ commonly used in the literature (see Appendix A).

[9] We implement the test as a bootstrap overidentification test. Let $\hat{r}_j = (\hat{q}_0(p_j) - \hat{q}_1(p_j)) - \tau\hat{\epsilon} - \hat{\eta}\pi(\hat{q}_0(p_j))$ be the residuals from (9) and let $J_0 \equiv \sum_{j=1}^{n} \hat{r}_j^2$. For each bootstrap draw $b = 1, \ldots, N_B$, we re-estimate the quantile functions, re-center $\tilde{y}_j^{(b)} = (\hat{q}_0^{(b)}(p_j) - \hat{q}_1^{(b)}(p_j)) - \hat{r}_j$, and obtain the bootstrapped residuals $\{\hat{r}_j^{(b)}\}_j$

The test rejects at the 1% or lower level when using data at or beyond the 17th quantile.

4 Identification definitions

In the remainder of the paper, we consider more complicated settings in which ϵ and η are random variables. These cases generally lead to partial identification. To lay the groundwork for this analysis, we begin in this section by formally defining the identification problem and the relevant objects.

The primitive unobservable is the joint distribution of (ϵ, η, v_0). Let F be such a distribution. The model implies that $\epsilon \geq 0$ and $\eta \in [-1, 0]$, so we restrict our attention to the set of distributions $\mathcal{F} \equiv \{F : \mathbb{P}_F[\epsilon \geq 0, \eta \in [-1, 0]] = 1\}$, where $\mathbb{P}_F$ denotes probability taken when (ϵ, η, v_0) is distributed like F. In some of our results, we further restrict F to lie in a subset $\mathcal{F}^\dagger \subseteq \mathcal{F}$ that reflects additional assumptions placed on F.

Each $F \in \mathcal{F}$ directly implies a distribution of system 0 virtual earnings through its marginal distribution of v_0. It also implies a distribution of system 1 virtual earnings $\nu(v_0, \epsilon, \eta)$, where ν is defined in (VE). The identified set, $\mathcal{F}^\star$, is defined as the subset of $\mathcal{F}^\dagger$ that matches the distributions of G_0 and G_1 up to some pre-specified points:

$$\mathcal{F}^\star \equiv \Big\{F \in \mathcal{F}^\dagger : \mathbb{P}_F[v_0 \leq y] = G_0(y) \text{ for all } \bar{y} < y < \bar{\bar{y}}_0, \\ \mathbb{P}_F[\nu(v_0, \epsilon, \eta) \leq y] = G_1(y) \text{ for all } \bar{y} \leq y < \bar{\bar{y}}_1\Big\}, \tag{ID}$$

where $\bar{\bar{y}}_0$ and $\bar{\bar{y}}_1$ are known and specified by the researcher.

In Sections 5 and 6, where we only consider top brackets, we set $\bar{\bar{y}}_0 = \bar{\bar{y}}_1 = \infty$. We record this condition for future reference.

Condition T. (Top bracket data) $\bar{\bar{y}}_0 = \bar{\bar{y}}_1 = \infty$.

Under Condition T, any $F \in \mathcal{F}^\star$ must have marginal distribution of v_0 equal to G_0 and must imply a marginal distribution of $v_1 = \nu(v_0, \epsilon, \eta)$ equal to G_1. We relax Condition T in Section 7, where we consider interior brackets and top brackets but with limited extrapolation.

Each $F \in \mathcal{F}$ generates a value for a lower-dimensional target parameter of interest, such as the average compensated elasticity, $\mathbb{E}_F[\epsilon]$. Our goal is to characterize all values that such a target parameter can take as F ranges across its identified set $\mathcal{F}^\star$. We say that a bound on a target parameter is valid if it is satisfied for all $F \in \mathcal{F}^\star$. We say

from regressing $\tilde{y}_j^{(b)}$ onto τ and $\pi(\hat{q}_0^{(b)}(p_j))$. The bootstrap test statistic is $J^{(b)} \equiv \sum_{j=1}^{n} (\hat{r}_j^{(b)})^2$. The p-value is $N_B^{-1} \sum_{b=1}^{N_B} \mathbb{1}[J^{(b)} \geq J_0]$.

that a finite bound is sharp if it can be obtained by some $F \in \mathcal{F}^\star$. We say that an infinite bound is sharp if any finite bound can be obtained by some $F \in \mathcal{F}^\star$.

In order to derive a sharp bound, we need to first understand when $\mathcal{F}^\star$ is non-empty, so that the problem is well-posed. We say that the model is misspecified if $\mathcal{F}^\star$ is empty. We say that a model is not misspecified if $\mathcal{F}^\star$ is non-empty. We refer to conditions that characterize when $\mathcal{F}^\star$ is empty as testable implications. A testable implication is called sharp if it is both sufficient and necessary for $\mathcal{F}^\star$ to be empty.

5 Heterogeneous substitution effects

In this section, we assume that η is deterministic and known, as it is, for example, under the common assumption of no income effects ($\eta = 0$). We record this as a condition for future reference.

Condition D. (Deterministic income effects) $\mathcal{F}^\dagger$ is the subset of $\mathcal{F}$ under which the distribution of η is degenerate and equal to a known constant.

We focus on identification of the means of ϵ, both unconditional and conditional on bunching status. Throughout this section, we maintain Condition T. Under this condition and Condition D, the only part of F that is unknown is the conditional distribution of ϵ given v_0.

5.1 Testable implications

We begin with a sharp testable implication that characterizes when $\mathcal{F}^\star$ is non-empty.

Proposition 4. Maintain Conditions T and D. Let $\zeta(y_0; \eta) \equiv y_0 - \eta\pi(y_0)$. The model is not misspecified if and only if $\mathbb{P}[\zeta(y_0; \eta) \leq y] \leq G_1(y)$ for all $y \geq \bar{y}$.

The quantity $\zeta(y_0; \eta)$ can be thought of as the "zero-elasticity earnings response" (ZER) that would occur in system 1 if compensated elasticities were zero. If income effects were also zero, then individuals would not change earnings, so $\zeta(y_0; 0) = y_0$.

Proposition 4 shows that whether the model is misspecified can be determined by comparing the ZER distribution to the actual system 1 earnings distribution, G_1. If the model is not misspecified, then G_1 can be produced by some distribution of compensated elasticities. This distribution is dominated by the ZER distribution because compensated elasticities are non-negative. The converse direction is shown constructively: if the ZER distribution dominates G_1, then it is possible to construct a distribution of ϵ that is non-negative and reproduces G_1, implying that the model is not misspecified.

The testable implication in Proposition 4 can be equivalently stated in terms of the quantile functions of the ZER and the actual system 1 earnings distribution: $\zeta(q_0(p);\eta) \geq q_1(p)$ for all p. Appendix Figure C.3 reports estimates of the difference between these two quantile functions when taking $\eta = 0$. The difference is larger than zero for all values of p, so the testable implication is satisfied. One-sided tests do not reject the null that the difference is non-negative for any p at any conventional significance level. The testable implication in Proposition 4 is sharp (sufficient and necessary), so this provides strong evidence that the model is not misspecified once one allows for heterogeneous compensated elasticities.

5.2 Trimming bounds

Solving (VE) for ϵ shows how the compensated elasticity is related to the change in virtual earnings:

$$\epsilon = \frac{1}{\tau}\left(v_0 - \eta\pi(v_0) - v_1\right) \equiv \frac{1}{\tau}\left(\zeta(v_0;\eta) - v_1\right). \tag{CE}$$

This relationship shows that the compensated elasticity is the (scaled) difference between the ZER, $\zeta(v_0;\eta)$, and the virtual system 1 earnings, v_1. Condition T implies that $v_0 = y_0$ and that $v_1 = y_1$ for all but the bunchers, making only η and ϵ unknown. In Section 3, we used (CE) with a known η to directly identify a deterministic ϵ.

Allowing ϵ to be random creates two challenges for identification, both of which can be seen from (CE). First, even if virtual earnings were fully observed as $y_1 = v_1$ and $y_0 = v_0$, the marginal distributions of y_0 and y_1 would be compatible with multiple joint distributions of (y_0, y_1), and therefore also multiple distributions of ϵ. Second, virtual earnings are not fully observed because they are censored for the bunchers: (BU) shows that individuals with $y_1 = \bar{y}$ have virtual earnings $v_1 \leq \bar{y}$.

These challenges mean that the average compensated elasticity will not in general be point identified for either the bunchers, the non-bunchers, or the overall population. There is, however, still structure for deriving informative bounds. This structure comes from two sources.

First, there is the probabilistic relationship that the system 0 distribution is a mixture of the earnings distributions of bunchers and non-bunchers. That is,

$$G_0(y) = \underbrace{G_{0|0}(y)(1-\bar{p})}_{\text{non-bunchers}} + \underbrace{G_{0|1}(y)\bar{p}}_{\text{bunchers}}, \tag{10}$$

where we use the shorthand notation $G_{0|b}(y) \equiv \mathbb{P}[y_0 \leq y | B = b]$ with $B \equiv \mathbb{1}[v_1 \leq \bar{y}] =$

$\mathbb{1}[y_1 = \bar{y}]$ denoting bunching status. While G_0 is observable, neither $G_{0|0}$ nor $G_{0|1}$ is observable, because B depends on y_1 and we do not observe the joint distribution of (y_0, y_1). However, because the proportion of bunchers $\bar{p}$ is observable, $G_{0|0}$ and $G_{0|1}$ can be bounded in terms of the $\bar{p}$ and $(1-\bar{p})$ quantiles of G_0 by using a trimming argument similar to that developed by Horowitz and Manski (1995) and applied to program evaluation by Lee (2009).

The second source of structure is the model implication that the ZER is weakly larger than actual system 1 earnings: $\zeta(v_0; \eta) \geq v_1$. This inequality must be true with probability one for ϵ to be non-negative with probability one, because $\tau > 0$; see (CE). If $\eta = 0$, then this inequality reduces to $v_0 \geq v_1$, reflecting the observation that virtual earnings under system 1 must be unambiguously lower than system 0 earnings in the absence of income effects. More generally, the inequality implies that the distribution of $\zeta(v_0; \eta)$ must dominate the distribution of v_1 for the non-bunchers.

Exploiting both sources of structure leads to the following result, which provides sharp bounds on average compensated elasticities, both unconditional and conditional on buncher status.

Proposition 5. Maintain Conditions T and D. The following bounds are valid:

$$\mathbb{E}[\epsilon] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta)] - \mathbb{E}[y_1]\right) \tag{P5-LB}$$

$$\mathbb{E}[\epsilon | B = 1] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta) | y_0 \leq q_0(\bar{p})] - \bar{y}\right) \tag{P5-LB-B1}$$

$$\mathbb{E}[\epsilon | B = 0] \geq \frac{1}{\tau}\left(\int \zeta(y;\eta)\, d\overline{G}(y) - \mathbb{E}[y_1 | y_1 > \bar{y}]\right), \tag{P5-LB-B0}$$

$$\text{where} \quad \overline{G}(y) \equiv \min\left\{G_0(y)/(1-\bar{p}), G_{1|0}(\zeta(y;\eta))\right\},$$

$$\text{and} \quad \mathbb{E}[\epsilon | B = 0] \leq \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0;\eta) | y_0 > q_0(\bar{p})\right] - \mathbb{E}[y_1 | y_1 > \bar{y}]\right). \tag{P5-UB-B0}$$

If the model is not misspecified, then (P5-LB-B1) and (P5-UB-B0) are sharp, while the sharp upper bounds on $\mathbb{E}[\epsilon]$ and $\mathbb{E}[\epsilon | B = 1]$ are infinite. The bounds (P5-LB) and (P5-LB-B0) are sharp under a "single-crossing" condition that $G'_{1|0}(y)$ and $G'_0(y)/(1-\bar{p})$ only cross a single time as functions of y, where $G'_{1|0}$ and G'_0 are first derivatives (densities).

The first three rows of Table 2 report estimates of the bounds in Proposition 5 in our application under the assumption of no income effects ($\eta = 0$). The bounds on the non-bunchers' average compensated elasticity are remarkably tight. The sharp upper bound is obtained when the entirety of the $\bar{p}$ mass of bunchers corresponds to the left $\bar{p}$ tail of the distribution of $\zeta(y_0; 0) = y_0$. In this scenario, the bunchers would need to have an average compensated elasticity of at least 0.039, which is their sharp lower bound reported in Table 2. At the same time, the largest average compensated elasticity that the non-bunchers could have while still rationalizing the remaining differences between the two distributions is 0.064.

The lower bound for the non-bunchers is more subtle. It is obtained by using the bunchers to explain as much of the difference between the distributions of y_1 and $y_0 = \zeta(y_0; 0)$ as possible. If the bunching moment $\bar{p}$ is sufficiently large to explain all of the downward shift from y_0 to y_1, then it is also possible that all of this downward shift was created by bunchers moving to the kink. The non-bunchers would have $y_0 = y_1$ implying that

$$G_{1|0}(y) = G_{0|0}(y) = \frac{\mathbb{P}[y_0 \leq y, B = 0]}{1 - \bar{p}} \leq \frac{G_0(y)}{1 - \bar{p}},$$

so that $\bar{G}(y) = G_{1|0}(y)$ for all y, rendering the lower bound (P5-LB-B0) zero. This is what happens in our application. While it leads to a sharp lower bound of zero for the non-bunchers, it also leads to a sharp lower bound of 0.064 for the combined group of top-earners. This combined lower bound could be entirely driven by the bunchers, who would need to be highly elastic to explain all of the differences between the system 0 and system 1 distributions.[10]

In more general cases, it might not be possible to rationalize the differences between the distributions of y_1 and y_0 (or $\zeta(y_0; \eta)$, more generally) as only arising from the bunchers. The lower bound for the non-bunchers and for the combined group of top-earners would then be obtained by placing the $\bar{p}$ mass of bunchers in the distribution of y_0 in a way that minimizes the remaining differences between the distributions of y_0 and y_1. The lower bound (P5-LB-B0) represents one way to do this, but it is only sharp if it can be attained by a non-negative distribution of ϵ. In Appendix SA.4, we show that this is possible if the single-crossing condition holds. We also derive an expression for the sharp lower bound when the single-crossing condition is relaxed to allow for any finite number of crossings.

5.3 Prior upper bounds

The sharp upper bound for bunchers in Proposition 5 is infinite because the bunchers could be arbitrarily elastic and they would still locate at the kink in system 1. This also leads the sharp upper bound on the overall average elasticity to be infinite. This conclusion depends on extreme behavior and seems overly conservative. We can dis-

[10]The lower bound for the combined top-earners and the upper bound for the non-bunchers are coincidentally equal up to three decimal points, but not beyond that.

Table 2: Compensated elasticities with heterogeneity and no income effects

	ϵ_{MAX}	Estimate	95% CI
Bunchers	∞	$[0.039, \infty]$	$(0.039, \infty)$
Non-bunchers	∞	$[0.000, 0.064]$	$(0.000, 0.064)$
All	∞	$[0.064, \infty]$	$(0.063, \infty)$
	3	$[0.064, 0.092]$	$(0.063, 0.092)$
	1.5	$[0.064, 0.078]$	$(0.063, 0.078)$
	1	$[0.064, 0.073]$	$(0.063, 0.073)$

Notes: This table shows estimates of the sharp bounds for average compensated elasticities when allowing for heterogeneous substitution effects and assuming no income effects. The first three rows impose no additional assumptions. The last three rows maintain different choices of ϵ_{MAX}, which puts an upper bound on the average compensated elasticity of bunchers. Confidence intervals are constructed using the method of Imbens and Manski (2004) with the bootstrap to account for estimating the pre-kink distribution.

cipline it by placing a prior upper bound on the average compensated elasticity for bunchers. This prior upper bound then leads to a non-trivial upper bound on the average elasticity for the combined population of top-earners.

Condition DB. (Deterministic income effects and bounded compensated elasticities for the bunchers) $\mathcal{F}^{\dagger}$ is the subset of $\mathcal{F}$ under which (i) the distribution of η is degenerate and equal to a known constant and (ii) $\mathbb{E}_F[\epsilon|B = 1] = \mathbb{E}_F[\epsilon|\nu(v_0, \epsilon, \eta) \leq \bar{y}] \leq \epsilon_{\text{MAX}}$, where ϵ_{MAX} is a known constant.

Proposition 6. Maintain Conditions T and DB. The following bound is valid:

$$\mathbb{E}[\epsilon] \leq \epsilon_{\text{MAX}}\bar{p} + \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0; \eta)|y_0 > q_0(\bar{p})\right] - \mathbb{E}[y_1|y_1 > \bar{y}]\right)(1 - \bar{p}). \tag{11}$$

If the model is not misspecified, then the sharp lower bound on $\mathbb{E}[\epsilon]$ is given by (P5-LB), and the sharp upper bound is given by (11).

The last three rows of Table 2 report estimated bounds on the average compensated elasticity for all top-earners under three different values of ϵ_{MAX}. Figure C.4 in Appendix C shows that fewer than ten percent of the studies considered in the meta-analysis conducted by Neisser (2021) find elasticities above 1. Fewer than 25% find elasticities above 1 for the self-employed and elasticities in Nordic countries tend to be much smaller than in other countries. This suggests that even our most aggressive choice of $\epsilon_{\text{MAX}} = 1$ is conservative.[11] This choice produces a narrow and tightly estimated bound on average compensated elasticities of $[0.064, 0.073]$. Taking $\epsilon_{\text{MAX}} = 1.5$, which is slightly above the largest estimate for the self-employed reported by Neisser (2021), produces bounds that are only marginally wider. Even taking $\epsilon_{\text{MAX}} = 3$ produces an informative upper bound of 0.092.

6 Heterogeneous income and substitution effects

In this section, we allow η to be a random variable, so that both compensated elasticities and income effects are heterogeneous. This means that instead of Condition D, we take $\mathcal{F}^\dagger = \mathcal{F}$. We focus on identification of the means of η, ϵ, and $\epsilon^u = \epsilon + \eta$, both unconditional and conditional on bunching status. We continue to maintain Condition T, which we relax in the next section.

6.1 Testable implications

The following proposition shows that the sharp testable implications when $\mathcal{F}^\dagger = \mathcal{F}$ are similar to those derived in Proposition 4 under Condition D. The difference is that instead of taking η to be the value specified in Condition D, we take it to be as large (in magnitude) as it can be by setting $\eta = -1$.

Proposition 7. Maintain Condition T and $\mathcal{F}^\dagger = \mathcal{F}$. The model is not misspecified if and only if $\mathbb{P}[\zeta(y_0; -1) \leq y] \leq G_1(y)$ for all $y \geq \bar{y}$.

Proposition 7 follows from the observation that the ZER, $\zeta(y_0; \eta) \equiv y_0 - \eta\pi(y_0)$, is a decreasing function of η. This implies that the stochastic dominance condition given in Proposition 4 for homogeneous income effects is weakest when $\eta = -1$. If that dominance condition is satisfied with $\eta = -1$, then the model with homogeneous income effects set to $\eta = -1$ is not misspecified, implying that a model with heterogeneous income effects also cannot be misspecified. Conversely, if a model with heterogeneous income effects is not misspecified, then we know that its ZER—which is random due to both y_0 and η—still cannot be larger than when $\eta = -1$ deterministically.

In our application, we found strong evidence that the model with heterogeneous elasticities but no income effects ($\eta = 0$) was not misspecified (Appendix Figure C.3). We conclude that the less restrictive model with heterogeneous income effects is also not misspecified.

[11] This choice is also well above the estimated lower bound (P5-LB-B1) of 0.039. Note also that ϵ_{MAX} is a bound on the *average* compensated elasticity of ϵ for bunchers, so it still allows some bunchers to have $\epsilon > \epsilon_{\text{MAX}}$.

6.2 Income effects

Proposition 7 implies that if the model is not misspecified, then a distribution of η that is degenerate at $\eta = -1$ is consistent with the data. As a consequence, the sharp lower bound on the expectation of η is -1. Because $\zeta(y_0; \eta)$ is increasing in η, there is also a maximal value of η under which the deterministic dominance condition in Proposition 4 is satisfied. We denote this value as

$$\bar{\eta} \equiv \max \left\{ \eta \in [-1, 0] : \mathbb{P}[\zeta(y_0; \eta) \leq y] \leq G_1(y) \text{ for all } y \geq \bar{y} \right\},$$

which is well-defined as long as the model with heterogeneous η is not misspecified (Proposition 7). It follows that any distribution of η supported on $[-1, \bar{\eta}]$ is consistent with the data, which implies that the sharp identified set for average income effects must contain $[-1, \bar{\eta}]$.

Proposition 8. Maintain Condition T and $\mathcal{F}^\dagger = \mathcal{F}$. Suppose that the model is not misspecified. Then the sharp identified sets for $\mathbb{E}[\eta]$, $\mathbb{E}[\eta | B = 0]$, and $\mathbb{E}[\eta | B = 1]$ contain $[-1, \bar{\eta}]$. In particular, the sharp lower bound is always -1.

In our application, we did not reject the model with ϵ heterogeneous and $\eta = 0$. This means that $\bar{\eta} = 0$. Because η must be non-positive, we conclude that the *sharp* identified set for the distribution of income effects is completely uninformative, at least in our application.

6.3 Compensated elasticities

Allowing for heterogeneous income effects changes the conclusions we can draw about compensated elasticities. This is because there is a trade-off between income and substitution effects: larger values of ϵ produce larger virtual earnings responses, while smaller (more negative) values of η produce smaller virtual earnings responses. The implication for identification is that larger compensated elasticities can be rationalized by stronger income effects. However, the adding-up condition represented by Engel aggregation places a limit on the extent to which income effects can rationalize larger compensated elasticities. The following proposition characterizes those limits.

Proposition 9. Maintain Condition T and set $\mathcal{F}^\dagger = \mathcal{F}$. The following bounds are valid:

$$\mathbb{E}[\epsilon] \geq \frac{1}{\tau}\left(\mathbb{E}[y_0] - \mathbb{E}[y_1]\right) \tag{P9-LB}$$

$$\mathbb{E}[\epsilon|B=1] \geq \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \leq q_0(\bar{p})] - \bar{y}\right) \tag{P9-LB-B1}$$

$$\mathbb{E}[\epsilon|B=0] \geq \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \leq q_0(1-\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right), \tag{P9-LB-B0}$$

$$\text{and} \quad \mathbb{E}[\epsilon|B=0] \leq \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0;-1)|y_0 > q_0(\bar{p})\right] - \mathbb{E}[y_1|y_1 > \bar{y}]\right). \tag{P9-UB-B0}$$

If the model is not misspecified, then (P9-UB-B0) is sharp and the sharp upper bounds on $\mathbb{E}[\epsilon]$ and $\mathbb{E}[\epsilon|B=1]$ are infinite. The bounds (P9-LB) and (P9-LB-B1) are sharp if $\bar{\eta}=0$.

The first three rows of Table 3 report estimates of the bounds in Proposition 9 in our application. The lower bounds on the compensated elasticity are the same as they were under the assumption of no income effects (Table 2). Note that while the lower bound for the non-bunchers is not sharp in general, the sharp bound must be weakly smaller than the lower bound (P5-LB-B0) in Proposition 5 derived for deterministic η. The latter was zero in our application, so we conclude that the sharp lower bound for the non-bunchers under heterogeneous income effects is also zero. The upper bound for the non-bunchers increases considerably from 0.064 without income effects to 0.361 with heterogeneous income effects. This upper bound is sizable compared to other estimates from Nordic countries, although still relatively small among estimates for the self-employed (see Figure C.4).

6.4 Uncompensated elasticities

Adding η to both sides of (CE) gives an expression for the uncompensated elasticity:

$$\epsilon^u = \epsilon + \eta = \frac{1}{\tau}\left(\zeta(v_0;\eta) + \tau\eta - v_1\right) \equiv \frac{1}{\tau}(\zeta^u(v_0;\eta) - v_1). \tag{UE}$$

Whereas $\zeta(v_0;\eta)$ is a decreasing function of η, its uncompensated counterpart, $\zeta^u(v_0;\eta)$, is a strictly *increasing* function of η. This follows from Proposition 1 because $\zeta^u(v_0;\eta) \equiv v_0 + \eta(\tau - \pi(v_0))$ and $\pi(v_0) < \tau$. Accounting for this difference and applying an argument similar to Propositions 5 and 9 leads to the following bounds.

Proposition 10. Maintain Condition T and set $\mathcal{F}^\dagger = \mathcal{F}$. The following bounds are valid:

$$\mathbb{E}[\epsilon^u] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;-1)] - \tau - \mathbb{E}[y_1]\right) \tag{P10-LB}$$

$$\mathbb{E}[\epsilon^u|B=1] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;-1)|y_0 \leq q_0(\bar{p})] - \tau - \bar{y}\right) \tag{P10-LB-B1}$$

$$\mathbb{E}[\epsilon^u|B=0] \geq \frac{1}{\tau}\left(\int \zeta(y;-1)\, d\overline{G}(y) - \tau - \mathbb{E}[y_1|y_1 > \bar{y}]\right) \tag{P10-LB-B0}$$

$$\text{where} \quad \overline{G}(y) \equiv \min\left\{G_0(y)/(1-\bar{p}), G_{1|0}(\zeta(y;-1))\right\},$$

$$\text{and} \quad \mathbb{E}[\epsilon^u|B=0] \leq \frac{1}{\tau}\left(\mathbb{E}\left[y_0|y_0 > q_0(\bar{p})\right] - \mathbb{E}[y_1|y_1 > \bar{y}]\right). \tag{P10-UB-B0}$$

If the model is not misspecified, then (P10-LB) and (P10-LB-B1) are sharp, while the sharp upper bounds on $\mathbb{E}[\epsilon^u]$ and $\mathbb{E}[\epsilon^u|B = 1]$ are infinite. If the single-crossing condition in Proposition 5 is satisfied, then (P10-LB-B0) is sharp. If $\bar{\eta} = 0$, then (P10-UB-B0) is also sharp.

In Appendix SA.4, we extend Proposition 10 to characterize the sharp lower bound on $\mathbb{E}[\epsilon^u|B = 0]$ when the single-crossing condition is relaxed.

The first three rows of Table 3 report estimated bounds on uncompensated elasticities. These are estimates of the sharp bounds because in our data the single-crossing condition is satisfied and $\bar{\eta} = 0$. The upper bound on the average uncompensated elasticity among non-bunchers is positive, but small. By contrast, the most informative lower bound is -0.642 on the overall average, $\mathbb{E}[\epsilon^u]$, with subgroup lower bounds being less informative. While relatively wide, these bounds are sharp, so it is not possible to obtain stronger conclusions about uncompensated elasticities without imposing stronger assumptions.

6.5 Prior bounds

In Section 5.3, we disciplined the extreme upper bounds on compensated elasticities with a prior bound for the bunchers, represented as Condition DB. The following proposition extends those results to the case with heterogeneous income effects while also allowing for a prior lower bound on income effects.

Condition B. (Bounded elasticities) $\mathcal{F}^\dagger$ is the subset of $\mathcal{F}$ under which (i) $\mathbb{E}_F[\epsilon|B = 1] \leq \epsilon_{\text{MAX}}$, where ϵ_{MAX} is a known constant and (ii) $\mathbb{E}_F[\eta|v_0, B = b] \geq \eta_{\text{MIN}}$ for $b = 0, 1$ and almost every v_0, where $\eta_{\text{MIN}} \in [-1, 0]$ is a known constant.

Proposition 11. Maintain Conditions T and B. The following bounds are valid:

$$\mathbb{E}[\epsilon] \geq \frac{1}{\tau}\left(\mathbb{E}[y_0] - \mathbb{E}[y_1]\right) \tag{P11-LB-ϵ}$$

$$\mathbb{E}[\epsilon^u] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta_{\text{MIN}})] + \tau\eta_{\text{MIN}} - \mathbb{E}[y_1]\right) \tag{P11-LB-ϵ^u}$$

$$\mathbb{E}[\epsilon] \leq \epsilon_{\text{MAX}}\bar{p} + \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta_{\text{MIN}})|y_0 \geq q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right)(1-\bar{p}) \tag{P11-UB-ϵ}$$

$$\mathbb{E}[\epsilon^u] \leq \epsilon_{\text{MAX}}\bar{p} + \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \geq q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right)(1-\bar{p}). \tag{P11-UB-ϵ^u}$$

Let $\underline{\epsilon}(\eta) \equiv \tau^{-1}(\mathbb{E}[\zeta(y_0;\eta)|y_0 \leq q_0(\bar{p})] - \bar{y})$ be the lower bound (P5-LB-B1). If the model is not misspecified, then:

- (P11-LB-ϵ) is sharp if $\bar{\eta} = 0$.
- (P11-LB-ϵ^u) is sharp if $\bar{\eta} \geq \eta_{\text{MIN}}$ and $\epsilon_{\text{MAX}} \geq \underline{\epsilon}(\eta_{\text{MIN}})$.
- (P11-UB-ϵ) is sharp if $\bar{\eta} \geq \eta_{\text{MIN}}$ and $\epsilon_{\text{MAX}} \geq \underline{\epsilon}(\eta_{\text{MIN}})$.
- (P11-UB-ϵ^u) is sharp if $\bar{\eta} = 0$.

Table 3 also reports sharp bounds on compensated and uncompensated elasticities under various prior bounds.[12] Imposing even a conservative upper bound of $\epsilon_{\text{MAX}} = 3$ leads to informative upper bounds for the overall average of both the compensated and uncompensated elasticity among top-earners. More aggressive choices do not tighten these upper bounds by much. However, adding a lower bound on *average* income effects provides considerable additional information.[13] For example, if $\epsilon_{\text{MAX}} = 1$ and we allow for large income effects by setting $\eta_{\text{MIN}} = -0.3$, then average compensated elasticities are bounded between 0.064 and 0.160. The bounds on average uncompensated elasticities tighten to $[-0.148, 0.073]$, suggesting that uncompensated elasticities are close to zero, consistent with the macro trends in hours worked reported in Boppart and Krusell (2020).

7 Interior brackets and partially observed distributions

In this section, we extend our analysis to interior tax brackets, as in Saez's (2010) application to the Earned Income Tax Credit (EITC). The extension can also be used to derive bounds for top brackets when using less of the earnings distribution. Both

[12] The conditions for sharpness in Proposition 11 ensure that ϵ_{MAX} and η_{MIN} do not render the model misspecified, the same as for Proposition 6. These conditions are easily met in our application. We find strong evidence that $\bar{\eta} = 0$, as previously noted. With $\eta_{\text{MIN}} = -0.5$, the lower bound of (P5-LB-B1) is 0.080, well below $\epsilon_{\text{MAX}} = 1$, which is the most aggressive choice that we consider.

[13] The bounds still allow some individuals to have $\eta < \eta_{\text{MIN}}$ as long as the conditional *average* of η is larger than η_{MIN}.

Table 3: Elasticities with heterogeneity and income effects

	η_{MIN}	ϵ_{MAX}	Compensated elasticity		Uncompensated elasticity	
			Estimate	95% CI	Estimate	95% CI
Bunchers	-1	∞	$[0.039, \infty]$	$(0.039, \infty)$	$[-0.932, \infty]$	$(-0.932, \infty)$
Non-bunchers	-1	∞	$[0.000, 0.361]$	$(0.000, 0.361)$	$[-0.852, 0.064]$	$(-0.852, 0.064)$
All	-1	∞	$[0.064, \infty]$	$(0.063, \infty)$	$[-0.642, \infty]$	$(-0.642, \infty)$
	-1	3	$[0.064, 0.386]$	$(0.063, 0.386)$	$[-0.642, 0.092]$	$(-0.642, 0.092)$
	-1	1.5	$[0.064, 0.372]$	$(0.063, 0.372)$	$[-0.642, 0.078]$	$(-0.642, 0.078)$
	-1	1	$[0.064, 0.367]$	$(0.063, 0.367)$	$[-0.642, 0.073]$	$(-0.642, 0.073)$
	-0.5	3	$[0.064, 0.239]$	$(0.063, 0.239)$	$[-0.289, 0.092]$	$(-0.289, 0.092)$
	-0.5	1.5	$[0.064, 0.225]$	$(0.063, 0.225)$	$[-0.289, 0.078]$	$(-0.289, 0.078)$
	-0.5	1	$[0.064, 0.220]$	$(0.063, 0.220)$	$[-0.289, 0.073]$	$(-0.289, 0.073)$
	-0.3	3	$[0.064, 0.180]$	$(0.063, 0.180)$	$[-0.148, 0.092]$	$(-0.148, 0.092)$
	-0.3	1.5	$[0.064, 0.166]$	$(0.063, 0.166)$	$[-0.148, 0.078]$	$(-0.148, 0.078)$
	-0.3	1	$[0.064, 0.161]$	$(0.063, 0.161)$	$[-0.148, 0.073]$	$(-0.148, 0.073)$
	-0.1	3	$[0.064, 0.121]$	$(0.063, 0.122)$	$[-0.007, 0.092]$	$(-0.007, 0.092)$
	-0.1	1.5	$[0.064, 0.107]$	$(0.063, 0.107)$	$[-0.007, 0.078]$	$(-0.007, 0.078)$
	-0.1	1	$[0.064, 0.102]$	$(0.063, 0.103)$	$[-0.007, 0.073]$	$(-0.007, 0.073)$

Notes: This table shows estimates of the sharp bounds for average compensated and uncompensated elasticities for bunchers, non-bunchers, and all top-earners when allowing for heterogeneous income and substitution effects under different choices of η_{MIN} *and* ϵ_{MAX}*. Confidence intervals are constructed using the method of Imbens and Manski (2004) with the bootstrap to account for estimating the pre-kink distribution.*

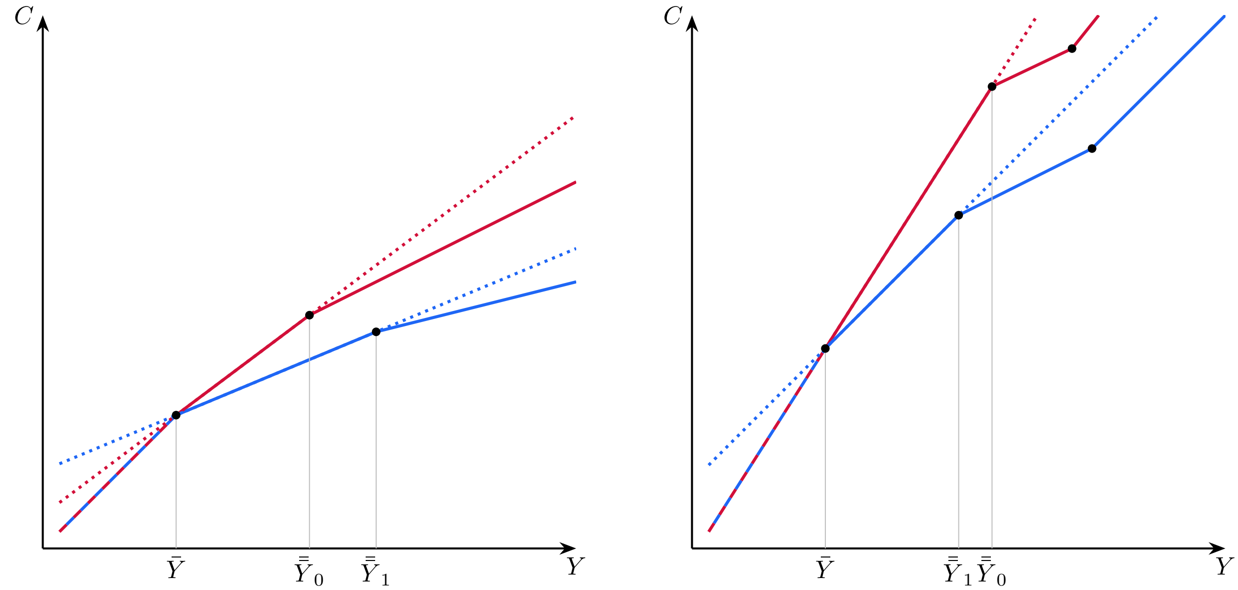
applications of this analysis amount to replacing Condition T to allow for some finite choices of $\bar{\bar{y}}_0$ and $\bar{\bar{y}}_1$ in the definition of (ID).

7.1 Interior brackets

We again consider a comparison between two tax systems, both of which have a bracket starting at $\bar{Y}$. In system 1, the bracket has marginal tax rate t_1 and ends at $\bar{\bar{Y}}_1$. In system 0, the bracket has marginal tax rate $t_0 < t_1$ and ends at $\bar{\bar{Y}}_0$. Below $\bar{Y}$, the two tax systems are the same, with a marginal tax rate weakly smaller than t_0. Above $\bar{\bar{Y}}_0$

Figure 3: Illustration of interior bracket reforms

(a) Convex (b) Non-convex



Notes: The system 0 budget set is shown in red and the system 1 budget set is shown in blue. The budget sets overlap to the left of $\bar{Y}$. The bracket starting at $\bar{Y}$ under system d has marginal tax rate t_d and ends at $\bar{\bar{Y}}_d$. Above $\bar{\bar{Y}}_d$, the tax systems can be convex (panel a) or non-convex (panel b), as long as the tax function is weakly larger than with the virtual linear tax system through $[\bar{Y}, \bar{\bar{Y}}_d]$ (depicted with the dotted lines).

and $\bar{\bar{Y}}_1$, the two tax systems could be either convex or non-convex, as long as the tax function is weakly larger than under the virtual linear tax systems through $[\bar{Y}, \bar{\bar{Y}}_0]$ and $[\bar{Y}, \bar{\bar{Y}}_1]$. This restriction ensures that the actual budget set for either system is a subset of the budget set under its virtual system.

Figure 3 illustrates both a convex and non-convex case. The convex case is a standard interior bracket in a progressive tax system. The non-convex case is similar to the EITC application studied by Saez (2010) in which transfers increase until $\bar{Y}$, remain constant until $\bar{\bar{Y}}_1$, and then phase out. A third case covered by these assumptions is a top bracket in which $\bar{\bar{Y}}_0$ and $\bar{\bar{Y}}_1$ are thresholds beyond which the researcher is unwilling to extrapolate; for this case, the tax functions above $\bar{\bar{Y}}_0$ and $\bar{\bar{Y}}_1$ are just equal to the respective virtual linear tax systems.

A hypothetical switch between systems 0 and 1 produces similar behavioral changes as in the top bracket case. Individuals with earnings at or below $\bar{Y}$ in system 0 will stay there, because their original bundle remains available under system 1. With convex preferences, no individual with earnings $Y_0 > \bar{Y}$ will relocate below $\bar{Y}$.

The behavior of the other individuals—those with $Y_0 > \bar{Y}$—can be partially characterized by their virtual earnings V_0 and V_1 under the virtual linear tax systems through $[\bar{Y}, \bar{\bar{Y}}_d]$. Those with $V_d \in (\bar{Y}, \bar{\bar{Y}}_d)$ will choose $Y_d = V_d$, because preferences are strictly convex and we have assumed that the virtual tax systems are more generous than the actual tax systems above $\bar{\bar{Y}}_d$.[14] While we restrict focus to those strictly above the kink under system 0 ($Y_0 > \bar{Y}$), some of these individuals may have $V_1 \leq \bar{Y}$, so they will choose to bunch at $\bar{Y}$ under system 1, again following Hall's (1973) argument. Under either system, those with $Y_d \geq \bar{\bar{Y}}_d$ must also have $V_d \geq \bar{\bar{Y}}_d$, because of convex preferences and our assumption that the virtual linear tax system is weakly more attractive at all earnings levels than the actual tax system.

These observations mean that we observe virtual earnings choices as actual earnings choices for $Y_0 \in (\bar{Y}, \bar{\bar{Y}}_0)$ and $Y_1 \in (\bar{Y}, \bar{\bar{Y}}_1)$. We also know that $V_1 \leq \bar{Y}$ for the bunchers under system 1, who have $Y_1 = \bar{Y}$. From Proposition 1, we still know that V_1 and V_0 are related (in logs) through the equation (VE). The implication is that the distributions of Y_0 and Y_1 between $(\bar{Y}, \bar{\bar{Y}}_0)$ and $[\bar{Y}, \bar{\bar{Y}}_1)$ can still be used to learn about the distributions of V_0 and V_1. Combining this knowledge with (VE) provides partial identification of features of the distribution of ϵ and η.

7.2 The distributions of earnings

We assume that the researcher continues to know the distribution $G_0(y)$ of y_0, which we continue to view as the distribution of y_0 *conditional* on $y_0 > \bar{y}$. Previously, we had defined $G_1(y)$ to be the distribution of y_1 conditional on $y_1 \geq \bar{y}$, which in the top bracket case was the same conditioning event as $y_0 > \bar{y}$. The interior bracket case creates a distinction between these two conditioning events: in the case that there is a kink at $\bar{y}$ under system 0, some of those with $y_1 = \bar{y}$ could have also had $y_0 = \bar{y}$; see Figure 3. Because of this distinction, we now define $G_1(y)$ to be the distribution of y_1 conditional on $y_0 > \bar{y}$, instead of conditional on $y_1 \geq \bar{y}$.

Redefining G_1 in this way means that it now depends on the joint distribution of y_0 and y_1. However, it is still identified because $y_0 > \bar{y}$ implies that $y_1 \geq \bar{y}$ and $y_0 = \bar{y}$

[14] If it were the case that $V_d > \bar{\bar{Y}}_d$, then strictly convex preferences would imply that there exists another point on the segment connecting Y_d and V_d that is both feasible under the actual tax system and strictly preferred to Y_d, creating a contradiction.

implies that $y_1 = \bar{y}$:

$$G_1(y) \equiv \frac{\mathbb{P}[y_1 \leq y, y_0 > \bar{y}]}{\mathbb{P}[y_0 > \bar{y}]}$$
$$= \frac{\mathbb{P}[\bar{y} \leq y_1 \leq y] - \mathbb{P}[y_0 \leq \bar{y}, \bar{y} \leq y_1 \leq y]}{\mathbb{P}[y_0 > \bar{y}]}$$
$$= \frac{\mathbb{P}[\bar{y} \leq y_1 \leq y] - \mathbb{P}[y_0 = \bar{y}, y_1 = \bar{y}]}{\mathbb{P}[y_0 > \bar{y}]} = \frac{\mathbb{P}[\bar{y} \leq y_1 \leq y] - \mathbb{P}[y_0 = \bar{y}]}{\mathbb{P}[y_0 > \bar{y}]}. \tag{12}$$

The final expression depends only on the marginal distributions of y_0 and y_1, implying that $G_1(y)$ is point identified. Note that (12) reduces back to $G_1(y) = \mathbb{P}[y_1 \leq y | y_1 \geq \bar{y}]$ in the top bracket case where $\mathbb{P}[y_0 = \bar{y}] = 0$ and $\mathbb{P}[y_0 > \bar{y}] = \mathbb{P}[y_1 \geq \bar{y}]$. Going forward, all probabilities and expectations should still be interpreted as conditional on $y_0 > \bar{y}$, which we continue to leave implicit in the notation.

We continue to assume that G_0 and G_1 are strictly increasing and absolutely continuous on $(\bar{y}, \bar{\bar{y}}_0)$ and $(\bar{y}, \bar{\bar{y}}_1)$, noting that G_1 could still have a mass point at $y = \bar{y}$, which we continue to denote as $\bar{p} \equiv G_1(\bar{y})$. We will not make use of these distributions at or above their upper thresholds, $\bar{\bar{y}}_d$, but there could be mass points at that location as well if there is bunching behavior above the upper threshold, as would be expected for example in Figure 3a. Note that after forming G_1 through (12), one can directly apply our identification results for the homogeneous case (Propositions 2 and 3) using quantiles that lie in $y_0 \in (\bar{y}, \bar{\bar{y}}_0)$ for system 0 and in $y_1 \in [\bar{y}, \bar{\bar{y}}_1)$ for system 1.[15]

7.3 Testable implications

As before, we first determine when the model is misspecified. The following proposition shows that the sharp testable implications are similar to those derived in Proposition 7, but on a restricted range of earnings.

Proposition 12. The model is not misspecified if and only if $\mathbb{P}[\zeta(y_0; -1) \leq y] \leq G_1(y)$ for all y such that $\bar{y} < y < \min\{\zeta(\bar{\bar{y}}_0; -1), \bar{\bar{y}}_1\}$.

The intuition behind Proposition 12 is essentially the same as the intuition behind Proposition 7. The domination condition is the same, but now only involves a comparison on a restricted range of earnings that respects the upper thresholds. Showing that satisfaction of the domination condition on this restricted set implies that $\mathcal{F}^\star$ is non-empty requires an additional step of extrapolating the distributions of virtual earnings; see Lemma VX in Appendix SA.3. A result analogous to Proposition 12 holds under Condition D, just with $\eta = -1$ replaced by the assumed-to-be-known value of η.

[15]This requires the weak additional condition that $\bar{\bar{y}}_0 > q_0(\bar{p})$, so that the amount of mass at $\bar{y}$ under system 1 is smaller than the amount of mass between $(\bar{y}, \bar{\bar{y}}_0)$ in system 0.

7.4 Bounds

We focus on characterizing bounds for a target subpopulation with system 0 virtual earnings $v_0 \in (\bar{y}, y_0^\star] \equiv \mathcal{Y}_0^\star$, where $y_0^\star$ is chosen by the researcher. Let $\mathbb{E}^\star[\cdot] \equiv \mathbb{E}[\cdot | v_0 \in \mathcal{Y}_0^\star]$ be the expectation operator conditional on the target range.

We impose three restrictions on the researcher's choice of $y_0^\star$. First, we require $y_0^\star < \bar{\bar{y}}_0$, so that v_0 is observed as y_0 for the target subpopulation. Second, we require $y_0^\star > q_0(\bar{p})$, so that the target subpopulation contains more mass than the bunching mass. Third, we require $\zeta(y_0^\star; -1) < \bar{\bar{y}}_1$, so that the largest value of v_1 attainable by a member of the target population is smaller than the upper limit of system 1 data that we use.

As in the top bracket case, we will maintain a prior upper bound of ϵ_{MAX} on the average elasticity for the bunchers, now restricted to the target subpopulation. We also maintain a prior lower bound on income effects.

Condition B*. (Bounded elasticities) $\mathcal{F}^\dagger$ is the subset of $\mathcal{F}$ under which (i) $\mathbb{E}_F^\star[\epsilon | B = 1] \leq \epsilon_{\text{MAX}}$, where ϵ_{MAX} is a known constant that satisfies $\epsilon_{\text{MAX}} \geq \underline{\epsilon}^\star \equiv (\zeta(q_0(\bar{p}); -1) - \bar{y})/\tau$, and (ii) $\mathbb{E}_F[\eta | v_0, B = b] \geq \eta_{\text{MIN}}$ for $b = 0, 1$ and almost every $v_0 \in \mathcal{Y}_0^\star$, where $\eta_{\text{MIN}} \in [-1, 0]$ is a known constant.

Condition B* requires that ϵ_{MAX} is larger than $\underline{\epsilon}^\star$, which is the value of ϵ that would be point identified with homogeneous elasticities and $\eta = -1$; see (3) in Proposition 2. In this case, the marginal buncher—and therefore all bunchers—would have compensated elasticity $\underline{\epsilon}^\star$. Requiring $\epsilon_{\text{MAX}} \geq \underline{\epsilon}^\star$ is needed to not assume away the homogeneous case a priori. In our application, $\underline{\epsilon}^\star = 0.11$, which is well below any reasonable prior bound ϵ_{MAX} that one would want to choose.

Bounds for the interior bracket case follow the same logic as the top bracket case, but with one important complication. In the top bracket case, objects involving the unconditional distribution of y_1, such as $\bar{p}$ and $\mathbb{E}[y_1]$, were identified from the observed system 1 distribution. In the interior bracket case, we need to focus on the subpopulation with $v_0 \in \mathcal{Y}_0^\star$ in order to exploit the model's implied relationship between v_1 and v_0. However, the distribution of y_1 conditional on $v_0 \in \mathcal{Y}_0^\star$ is not observed.

The next proposition gets around this problem by using a nested argument. First, we derive bounds that are conditional on the target buncher share $\bar{p}^\star \equiv \mathbb{P}[v_1 \leq \bar{y} | \bar{y} < v_0 \leq y_0^\star]$. Second, we bound the target buncher share, $\bar{p}^\star$. Then, we combine both bounds.

Proposition 13. Maintain Condition B*. Let $y_1^\star \equiv \zeta(y_0^\star; -1)$. Let $p_d^\star \equiv G_d(y_d^\star)$ for

$d = 0, 1$. Then define $q_1^{\text{L}} \equiv q_1(p_1^\star - p_0^\star)$ and $q_1^{\text{U}} \equiv q_1(p_0^\star)$. The following bounds are valid:

$$\mathbb{E}^\star[\epsilon] \geq \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \leq y_0^\star] - \mathbb{E}[y_1|q_1^{\text{L}} < y_1 \leq y_1^\star]\right) \tag{P13-LB-ϵ}$$

$$\mathbb{E}^\star[\epsilon^u] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta_{\text{MIN}})|y_0 \leq y_0^\star] + \tau\eta_{\text{MIN}} - \mathbb{E}[y_1|q_1^{\text{L}} < y_1 \leq y_1^\star]\right) \tag{P13-LB-ϵ^u}$$

$$\mathbb{E}^\star[\epsilon] \leq \frac{\bar{p}\epsilon_{\text{MAX}}}{p_0^\star} + \frac{1}{\tau}\Big(\mathbb{E}[\zeta(y_0;\eta_{\text{MIN}})|q_0(\bar{p}) \leq y_0 \leq y_0^\star] - \mathbb{E}[y_1|\bar{y} < y_1 \leq q_1^{\text{U}}]\Big)\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right) \tag{P13-UB-ϵ}$$

$$\mathbb{E}^\star[\epsilon^u] \leq \frac{\bar{p}\epsilon_{\text{MAX}}}{p_0^\star} + \frac{1}{\tau}\Big(\mathbb{E}[y_0|q_0(\bar{p}) \leq y_0 \leq y_0^\star] - \mathbb{E}[y_1|\bar{y} < y_1 \leq q_1^{\text{U}}]\Big)\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right). \tag{P13-UB-ϵ^u}$$

Define $\bar{\eta}^\star$ as

$$\bar{\eta}^\star \equiv \max\left\{\eta \in [-1, 0] : \mathbb{P}[\zeta(y_0;\eta) \leq y] \leq G_1(y) \text{ for all } y \in (\bar{y}, \min\{\zeta(\bar{\bar{y}}_0; -1), \bar{\bar{y}}_1\})\right\},$$

and assume that the model is not misspecified, so that $\bar{\eta}^\star \geq -1$ is well-defined. Then (P13-UB-ϵ) is sharp if $\bar{\eta}^\star \geq \eta_{\text{MIN}}$, and (P13-UB-$\epsilon^u$) is sharp if $\bar{\eta}^\star = 0$.

If Condition D is maintained, so that η is deterministic, then the same bounds are valid with the following changes: replace $y_1^\star$ by $\zeta(y_0^\star;\eta)$, replace conditional expectations of $\zeta(y_0;\eta)$ with conditional expectations of y_0 (keeping the conditioning event the same), and replace conditional expectations of $\zeta(y_0;\eta_{\text{MIN}})$ with $\zeta(y_0;\eta)$ (keeping the conditioning event the same). The resulting upper bounds are sharp if the model is not misspecified, the condition for which is given in Proposition 12 with $\eta = -1$ replaced by η.

Proposition 13 provides upper bounds on the average compensated and uncompensated elasticities for the target subpopulation with $y_0 \in \mathcal{Y}_0^\star$. The bounds are sharp given knowledge of the distributions of y_0 and y_1 up to $\bar{\bar{y}}_0$ and $\bar{\bar{y}}_1$, even though they only use these distributions up to $y_0^\star < \bar{\bar{y}}_0$ and $y_1^\star < \bar{\bar{y}}_1$. The upper bounds are obtained when the entirety of the $\bar{p}$ mass of bunchers corresponds to the left $\bar{p}$ tail of the distribution of y_0. The remaining individuals in the target population are non-bunchers whose virtual earnings v_1 are distributed as the left tail of y_1 above the kink.

The lower bounds in the propositions are potentially non-sharp. In Appendix SA.5, we develop a computational approach that approximately computes the sharp bounds by discretizing the distribution of (ϵ, η). This allows us to assess the extent to which the analytic bounds are non-sharp. The approach could also prove useful when considering additional assumptions beyond those considered explicitly in this paper.

Table 4: Elasticities for top-earners with less data allowing for heterogeneity

		Compensated elasticity		Uncompensated elasticity	
η_{MIN}	ϵ_{MAX}	Estimate	95% CI	Estimate	95% CI
0	3	$[0.000, 0.356]$	$(0.000, 0.358)$		
0	1.5	$[0.000, 0.213]$	$(0.000, 0.214)$		
0	1	$[0.000, 0.166]$	$(0.000, 0.166)$		
-1	3	$[0.000, 0.402]$	$(0.000, 0.404)$	$[-1.000, 0.356]$	$(-1.000, 0.358)$
-1	1.5	$[0.000, 0.260]$	$(0.000, 0.260)$	$[-1.000, 0.213]$	$(-1.000, 0.214)$
-1	1	$[0.000, 0.212]$	$(0.000, 0.213)$	$[-1.000, 0.166]$	$(-1.000, 0.166)$
-0.5	3	$[0.000, 0.379]$	$(0.000, 0.381)$	$[-0.500, 0.356]$	$(-0.500, 0.358)$
-0.5	1.5	$[0.000, 0.236]$	$(0.000, 0.237)$	$[-0.500, 0.213]$	$(-0.500, 0.214)$
-0.5	1	$[0.000, 0.189]$	$(0.000, 0.190)$	$[-0.500, 0.166]$	$(-0.500, 0.166)$
-0.3	3	$[0.000, 0.370]$	$(0.000, 0.372)$	$[-0.300, 0.356]$	$(-0.300, 0.358)$
-0.3	1.5	$[0.000, 0.227]$	$(0.000, 0.228)$	$[-0.300, 0.213]$	$(-0.300, 0.214)$
-0.3	1	$[0.000, 0.180]$	$(0.000, 0.180)$	$[-0.300, 0.166]$	$(-0.300, 0.166)$
-0.1	3	$[0.000, 0.361]$	$(0.000, 0.362)$	$[-0.100, 0.356]$	$(-0.100, 0.358)$
-0.1	1.5	$[0.000, 0.218]$	$(0.000, 0.219)$	$[-0.100, 0.213]$	$(-0.100, 0.214)$
-0.1	1	$[0.000, 0.170]$	$(0.000, 0.171)$	$[-0.100, 0.166]$	$(-0.100, 0.166)$

Notes: This table reports bounds on average compensated and uncompensated elasticities for top-earners with y_0 below the 10th quantile when using knowledge of the distribution of y_0 only up until that quantile and using full knowledge of the distribution of y_1. The first three rows assume no income effects and vary ϵ_{MAX} for the upper bound for the average compensated elasticity of bunchers with y_0 below the 10th quantile. The remaining rows allow for income effects by also varying η_{MIN}. Confidence intervals are constructed using the method of Imbens and Manski (2004) with the bootstrap to account for estimating the pre-kink distribution.

8 Estimates using less data

Proposition 13 allows us to compute bounds for our top bracket application that use less of the earnings distribution, mitigating concerns about extrapolation. Table 4 reports estimates of these bounds with $\bar{\bar{y}}_0 \equiv q_0(.10)$ set to be the 10th quantile of the system 0 earnings distribution and $y_0^\star = \bar{\bar{y}}_0$. In our application, all of G_1 is observed, so we do not need to choose $\bar{\bar{y}}_1$. As Proposition 13 shows, the bounds do not use the distribution of G_1 beyond $\zeta(\bar{\bar{y}}_0; -1)$. Table C.1 reports alternative bounds that set $\bar{\bar{y}}_0 = q_0(.05)$ and $\bar{\bar{y}}_0 = q_0(.20)$.

The first three rows of Table 4 show results for different values of ϵ_{MAX} under the assumption of no income effects ($\eta_{\text{MIN}} = 0$). Imposing even a conservative upper bound of $\epsilon_{\text{MAX}} = 3$ leads to an informative upper bound of 0.356 for the compensated elasticity. When $\epsilon_{\text{MAX}} = 1$, the upper bound decreases to 0.166, meaning that compensated elasticities are at most moderate. The upper bound is more sensitive to the choice of ϵ_{MAX} than in Table 3, which used the entire earnings distribution. This is because the bunchers—the group whose elasticities are disciplined by ϵ_{MAX}—constitute a larger share of the target subpopulation when using less data.

The next three rows of Table 4 show results for the same values of ϵ_{MAX} when allowing for arbitrary income effects. The upper bounds on the compensated elasticities change only marginally compared to the case with no income effects. The reason is that the tax reform cannot induce large changes in average tax rates for the target subpopulation because they have system 0 earnings that are closer to the kink. This implies that income effects can account for no more than a small share of the virtual earnings response. This intuition also explains why imposing other choices of η_{MIN} in the remaining rows of Table 4 also has little impact on the bounds.

Proposition 13 shows that the upper bounds in Table 4 are sharp, which means that it is not possible to obtain stronger conclusions given the maintained assumptions. The lower bounds, which are zero throughout, may not be sharp. In Appendix SA.5 we report the results from our computational procedure, which delivers the sharp lower bounds up to approximation error caused by discretization. These sharp lower bounds are also zero, suggesting that the reported analytic bounds are sharp as well.

9 Conclusion

We have shown how to point and partially identify labor supply elasticities from kinked budget sets in a model with income effects and individual heterogeneity in the elasticities. We applied our identification results to a kink in the Norwegian tax system for the self-employed. We found that even under weak assumptions, the bounds are quite informative: the uncompensated elasticities are close to zero, and the compensated elasticities are sufficiently small to rule out a sizable excess burden of taxation. While our paper is focused on labor supply elasticities and taxation, it is straightforward to adapt our identification results to other settings with kinked budget sets, such as firm profits (Best et al., 2015), intertemporal savings decisions (Best et al., 2019), and inheritance and gift giving (Glogowsky, 2021).

References

Ashenfelter, O. and M. W. Plant (1990): "Nonparametric Estimates of the Labor-Supply Effects of Negative Income Tax Programs," *Journal of Labor Economics*, 8, S396–S415.

Auten, G. and R. Carroll (1999): "The Effect of Income Taxes on Household Income," *The Review of Economics and Statistics*, 81, 681–693.

Bastani, S. and H. Selin (2014): "Bunching and Non-Bunching at Kink Points of the Swedish Tax Schedule," *Journal of Public Economics*, 109, 36–49.

Bertanha, M., C. Caetano, H. Jales, and N. Seegert (2024): "Bunching Estimation Methods," in *Handbook of Labor, Human Resources and Population Economics*, ed. by K. F. Zimmermann, Cham: Springer International Publishing, 1–44.

Bertanha, M., A. H. McCallum, A. Payne, and N. Seegert (2022): "Bunching Estimation of Elasticities Using Stata," *The Stata Journal: Promoting communications on statistics and Stata*, 22, 597–624.

Bertanha, M., A. H. McCallum, and N. Seegert (2023): "Better Bunching, Nicer Notching," *Journal of Econometrics*, 237, 105512.

Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015): "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan," *Journal of Political Economy*, 123, 1311–1355.

Best, M. C., J. S. Cloyne, E. Ilzetzki, and H. J. Kleven (2019): "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches," *The Review of Economic Studies*, 87, 656–690.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *The American Economic Review*, 96, 988–1012.

Blomquist, S., W. K. Newey, A. Kumar, and C.-Y. Liang (2021): "On Bunching and Identification of the Taxable Income Elasticity," *Journal of Political Economy*, 129, 2320–2343.

Boppart, T. and P. Krusell (2020): "Labor Supply in the Past, Present, and Future: A Balanced-Growth Perspective," *Journal of Political Economy*, 128, 118–157.

Burtless, G. and J. A. Hausman (1978): "The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment," *Journal of Political Economy*, 86, 1103–1130.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017): "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries," *American Economic Review*, 107, 3917–46.

Chetty, R. (2012): "Bounds on elasticities with optimization frictions: a synthesis of micro and macro evidence on labor supply," *Econometrica*, 80, 969–1018.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011): "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records *," *The Quarterly Journal of Economics*, 126, 749–804.

Coles, J. L., E. Patel, N. Seegert, and M. Smith (2022): "How Do Firms Respond to Corporate Taxes?" *Journal of Accounting Research*, 60, 965–1006.

Embrechts, P. and M. Hofert (2013): "A Note on Generalized Inverses," *Mathematical Methods in Operations Research*, 77, 423–432–.

Feldstein, M. (1999): "Tax Avoidance and the Deadweight Loss of the Income Tax," *The Review of Economics and Statistics*, 81, 674–680.

Gelber, A. M., D. Jones, D. W. Sacks, and J. Song (2021): "Using Nonlinear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Earnings Test," *American Economic Journal: Applied Economics*, 13, 150–193.

Glogowsky, U. (2021): "Behavioral responses to inheritance and gift taxation: Evidence from Germany," *Journal of Public Economics*, 193, 104309.

Goff, L. (2024): "Treatment Effects in Bunching Designs: The Impact of Mandatory Overtime Pay on Hours," .

Golosov, M., M. Graber, M. Mogstad, and D. Novgorodsky (2024): "How Americans Respond to Idiosyncratic and Exogenous Changes in Household Wealth and Unearned Income*," *The Quarterly Journal of Economics*, 139, 1321–1395.

Gruber, J. and E. Saez (2002): "The elasticity of taxable income: evidence and implications," *Journal of public Economics*, 84, 1–32.

Hall, R. E. (1973): "Wages, income and hours of work in the US labor force," *Income Maintenance and Labor Supply: Econometric Studies (Rand McNally, Chicago)*.

Horowitz, J. L. and C. F. Manski (1995): "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281–302.

Hungerman, D. M. and M. Ottoni-Wilhelm (2021): "Impure Impact Giving: Theory and Evidence," *Journal of Political Economy*, 129, 1553–1614.

Imbens, G. W. and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Kimball, M. S. and M. D. Shapiro (2008): "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" Working Paper 14208, National Bureau of Economic Research.

Kleven, H. J. (2016): "Bunching," *Annual Review of Economics*, 8, 435–464.

Kleven, H. J. and E. A. Schultz (2014): "Estimating Taxable Income Responses Using Danish Tax Reforms," *American Economic Journal: Economic Policy*, 6, 271–301.

Kleven, H. J. and M. Waseem (2013): "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan *," *The Quarterly Journal of Economics*, 128, 669–723.

Kostøl, A. R. and A. S. Myhre (2021): "Labor Supply Responses to Learning the Tax and Benefit Schedule," *American Economic Review*, 111, 3733–66.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 76, 1071–1102.

Mortenson, J. A. and A. Whitten (2020): "Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule," *American Economic Journal: Economic Policy*, 12, 402–32.

Neisser, C. (2021): "The Elasticity of Taxable Income: A Meta-Regression Analysis," *The Economic Journal*, 131, 3365–3391.

Pollinger, S. (2025): "Kinks Know More: Estimating Intensive and Participation Margin Responses Using Nonlinear Budget Sets," *Working Paper*.

Saez, E. (2010): "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, 180–212.

Saez, E., J. Slemrod, and S. H. Giertz (2012): "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review," *Journal of Economic Literature*, 50, 3–50.

A Estimating counterfactual earnings distributions

In this appendix, we formalize the assumptions used by Chetty et al. (2011) to recover the earnings distribution under tax system 0 from the observed earnings distribution under tax system 1.[16] We first establish the general identification result and then describe our estimation procedure.

A.1 Identification

Let G_1^{U} denote the *unconditional* cumulative distribution of observed earnings (under the kinked tax system) Y_1, and G_0^{U} the *unconditional* counterfactual cumulative distribution function of Y_0. (These definitions are in contrast to G_0 and G_1 in the main text, which are conditional on $y_0 > \bar{y}$.) The following proposition formalizes the assumption used by Chetty et al. (2011) and shows how to use it to recover G_0^{U}.

Proposition A.14. Suppose that G_0^{U} is continuous at $\bar{Y}$ and satisfies

$$G_0^{\text{U}}(z) = \begin{cases} G_1^{\text{U}}(z), & \text{if } z < \bar{Y} \\ \beta_0 + \beta_1 G_1^{\text{U}}(z) & \text{if } z \geq \bar{Y}. \end{cases} \tag{13}$$

Let $\bar{p}^{\text{U}} \equiv G_1^{\text{U}}(\bar{Y}) - \lim_{z\uparrow\bar{Y}} G_1^{\text{U}}(z)$. Then

$$\beta_0 = \frac{-\bar{p}^{\text{U}}}{1 - G_1^{\text{U}}(\bar{Y})} \quad \text{and} \quad \beta_1 = \frac{1 - \lim_{z\uparrow\bar{Y}} G_1^{\text{U}}(z)}{1 - G_1^{\text{U}}(\bar{Y})}. \tag{14}$$

Proof of Proposition A.14. Because $\lim_{z\to\infty} G_0^{\text{U}}(z) = 1$, we must have $\beta_0 + \beta_1 = 1$. Using the continuity of G_0^{U} at $\bar{Y}$ then gives

$$0 = G_0^{\text{U}}(\bar{Y}) - \lim_{z\uparrow\bar{Y}} G_0^{\text{U}}(z) = \beta_0 + (1-\beta_0)G_1^{\text{U}}(\bar{Y}) - \lim_{z\uparrow\bar{Y}} G_1^{\text{U}}(z) = \beta_0(1 - G_1^{\text{U}}(\bar{Y})) + \bar{p}^{\text{U}}.$$

Rearranging provides the stated expression for β_0. The expression for β_1 then follows:

$$\beta_1 = 1 - \left(\frac{-\bar{p}^{\text{U}}}{1 - G_1^{\text{U}}(\bar{Y})}\right) = \frac{1 - G_1^{\text{U}}(\bar{Y}) + \bar{p}^{\text{U}}}{1 - G_1^{\text{U}}(\bar{Y})} = \frac{1 - \lim_{z\uparrow\bar{Y}} G_1^{\text{U}}(z)}{1 - G_1^{\text{U}}(\bar{Y})}.$$

Q.E.D.

The assumption in equation (13) implies and is implied by equation (15) in Chetty et al.

[16] Other approaches to recover the counterfactual distribution have also been considered. For example, Saez (2010) recovers G_0 by assuming the counterfactual density above the kink is uniform. Bertanha et al. (2023) uses maximum likelihood-based estimators to jointly estimate labor supply elasticities and counterfactual earnings under restricted heterogeneity of elasticities (see also Bertanha et al., 2024).

(2011), which states that for $Y > \bar{Y}$, the density under tax system 0 is proportional to the density under tax system 1.

A.2 Estimation

We estimate G_0^{U} and G_1^{U} using Proposition A.14 and an i.i.d. sample $Y_1, \ldots, Y_n$ from G_1^{U}. In the data, bunchers do not appear exactly at $\bar{Y}$. Instead, the excess mass is dispersed over a small region near the kink. Following Chetty et al. (2011) and Bertanha et al. (2023), we exclude earnings in a region $[\bar{Y} - \kappa_{\mathrm{IN}}, \bar{Y} + \kappa_{\mathrm{IN}}]$ around the kink to avoid contamination from bunchers.

To estimate the distributions in the excluded region, we follow Chetty et al. (2011) and approximate G_0^{U} by a Kth degree polynomial on a wider window $[\bar{Y} - \kappa_{\mathrm{OUT}}, \bar{Y} + \kappa_{\mathrm{OUT}}]$ with $\kappa_{\mathrm{OUT}} > \kappa_{\mathrm{IN}}$:

$$G_0^{\mathrm{U}}(z) = \sum_{k=0}^{K} \gamma_k \left(z - \bar{Y}\right)^k . \tag{15}$$

Combining (15) with (13) implies that $\lim_{z \uparrow \bar{Y}} G_1^{\mathrm{U}}(z) = G_0^{\mathrm{U}}(\bar{Y}) = \gamma_0$. From the conclusion of Proposition A.14 we then get

$$G_1^{\mathrm{U}}(z; \gamma, \bar{p}^{\mathrm{U}}) = \begin{cases} \sum_{k=0}^{K} \gamma_k \left(z - \bar{Y}\right)^k & \text{if } \bar{Y} - \kappa_{\mathrm{OUT}} \le z < \bar{Y}, \\ \frac{\bar{p}^{\mathrm{U}}}{1-\gamma_0} + \frac{1-\gamma_0-\bar{p}^{\mathrm{U}}}{1-\gamma_0} \left(\sum_{k=0}^{K} \gamma_k \left(z - \bar{Y}\right)^k \right) & \text{if } \bar{Y} \le z \le \bar{Y} + \kappa_{\mathrm{OUT}}, \end{cases}$$

where we've made the dependence on the parameters $(\gamma, \bar{p}^{\mathrm{U}})$ explicit in the notation. We estimate $(\gamma, \bar{p}^{\mathrm{U}})$ using nonlinear least squares with the regressand taken as the rank of each observation. In doing so, we only use data near the kink but exclude a donut in the region that defines the bunchers:

$$\begin{aligned} (\hat{\gamma}, \hat{\bar{p}}^{\mathrm{U}}) \equiv \arg\min_{\gamma, \bar{p}^{\mathrm{U}}} \sum_{i=1}^{n} \mathbb{1}\Big[Y_i \in [\bar{Y} - \kappa_{\mathrm{OUT}}, \bar{Y} - \kappa_{\mathrm{IN}}) \cup (\bar{Y} + \kappa_{\mathrm{IN}}, \bar{Y} + \kappa_{\mathrm{OUT}}]\Big] \\ \times \left(\frac{1}{n} \left[\sum_{j=1}^{n} \mathbb{1}[Y_j \le Y_i] \right] - G_1^{\mathrm{U}}(Y_i; \gamma, \bar{p}^{\mathrm{U}}) \right)^2 . \end{aligned} \tag{16}$$

The resulting estimators of G_1^{U} and G_0^{U} for $z \in [\bar{Y} - \kappa_{\mathrm{OUT}}, \bar{Y} + \kappa_{\mathrm{OUT}}]$ are then

$$\hat{G}_1^{\mathrm{U}}(z) = G_1^{\mathrm{U}}(z; \hat{\gamma}, \hat{\bar{p}}^{\mathrm{U}}) \quad \text{and} \quad \hat{G}_0^{\mathrm{U}}(z) = \sum_{k=0}^{K} \hat{\gamma}_k \left(z - \bar{Y}\right)^k .$$

We estimate G_1^{U} away from the kink using a more flexible local linear regression:

$$(\hat{\theta}_0(z), \hat{\theta}_1(z)) \equiv \underset{\theta_0,\theta_1}{\arg\min} \sum_{i=1}^{n} \mathbb{1}\Big[Y_i \notin [\bar{Y} - \kappa_{\text{OUT}}, \bar{Y} + \kappa_{\text{OUT}}]\Big] \times \mathcal{K}\left(\frac{Y_i - z}{h}\right) \times \left(\frac{1}{n}\left[\sum_{j=1}^{n} \mathbb{1}[Y_j \leq Y_i]\right] - \theta_0 - \theta_1 Y_i\right)^2, \tag{17}$$

where $\mathcal{K}$ is a kernel function and h is the bandwidth. Then we take

$$\hat{G}_1^{\text{U}}(z) = \hat{\theta}_0(z) + \hat{\theta}_1(z) \times z.$$

We use Proposition A.14 to form an estimator for G_0^{U}:

$$\hat{G}_0^{\text{U}}(z) = \frac{-\hat{\bar{p}}^{\text{U}}}{1 - \hat{\gamma}_0 - \hat{\bar{p}}^{\text{U}}} + \frac{1 - \hat{\gamma}_0}{1 - \hat{\gamma}_0 - \hat{\bar{p}}^{\text{U}}} \left(\hat{\theta}_0(z) + \hat{\theta}_1(z) \times z\right), \tag{18}$$

where $\hat{\gamma}_0$ and $\hat{\bar{p}}^{\text{U}}$ were obtained from (16).

A.3 Implementation

In our application, we follow Chetty et al. (2011) and take $K = 8$, so that the counterfactual density $\hat{g}_0^{\text{U}}$ follows a 7th-degree polynomial. We set $\kappa_{\text{IN}} = 30{,}000$ NOK and $\kappa_{\text{OUT}} = 100{,}000$ NOK. The density $\hat{g}_0^{\text{U}}$ within this window is shown in Figure 2. To obtain the distributions for $z > \bar{Y} + \kappa_{\text{OUT}}$, we estimate (17) using an Epanechnikov kernel with the rule-of-thumb bandwidth. We conduct inference by bootstrapping the entire estimation procedure, resampling $Y_1, \ldots, Y_n$ with replacement.

B Proofs for Sections 2 and 3

Proof of Proposition 1. We start by showing how the parameters of the earnings function relate to income and substitution effects. Differentiation of the earnings function (EF) with respect to $1 - t$ yields

$$\frac{1}{Y^\star(1-t, R)} \frac{\partial Y^\star(1-t, R)}{\partial(1-t)} = \frac{\beta_t}{1-t}.$$

Rearranging gives the uncompensated elasticity:

$$\epsilon^u \equiv \frac{\partial Y^\star(1-t, R)}{\partial(1-t)} \frac{1-t}{Y^\star(1-t, R)} = \beta_t,$$

which we note is independent of both t and R, so applies to both system 0 and system 1. Let $V_0 \equiv Y^\star(1-t_0, R_0)$ and $V_1 \equiv Y^\star(1-t_1, R_1)$, so that $v_0 \equiv \log(V_0)$ and $v_1 \equiv \log(V_1)$. Then

$$v_1 - v_0 = -\epsilon^u \tau + \beta_R(\phi(R_1) - \phi(R_0)), \tag{19}$$

with $\tau \equiv \log(1-t_0) - \log(1-t_1)$. An expression for η can be found by differentiating (EF) evaluated at system 0:

$$\eta \equiv (1-t_0)\frac{\partial Y^\star(1-t_0, R_0)}{\partial R} = (1-t_0)V_0\beta_R\phi'(R_0).$$

Solving for β_R and substituting into (19) gives

$$v_1 - v_0 = -\epsilon^u\tau + \frac{\eta\,(\phi(R_1) - \phi(R_0))}{(1-t_0)V_0\phi'(R_0)} \equiv -\epsilon^u\tau - \eta(\pi(v_0) - \tau).$$

Substituting $\epsilon^u = \epsilon + \eta$ and rearranging establishes (VE).

The bound $\pi(v_0) < \tau$ follows immediately from ϕ being a strictly increasing function. To see that $\pi(v_0)$ is non-negative, write $\pi(v_0; t_1)$ as a function of t_1, expanding the definitions of τ and R_1:

$$\pi(v_0; t_1) \equiv \log(1-t_0) - \log(1-t_1) - \frac{\phi(R_0 + \bar{Y}(t_1 - t_0)) - \phi(R_0)}{(1-t_0)\exp(v_0)\phi'(R_0)}.$$

Differentiating with respect to t_1 at $v_0 > \bar{y}$ gives

$$\begin{aligned}\frac{\partial \pi(v_0; t_1)}{\partial t_1} &= \frac{1}{(1-t_1)} - \frac{\phi'(R_0 + (t_1 - t_0)\bar{Y})\bar{Y}}{(1-t_0)V_0\phi'(R_0)} \\ &> \frac{1}{(1-t_0)}\left(1 - \frac{\bar{Y}}{V_0}\right)\left(\frac{\phi'(R_0 + (t_1 - t_0)\bar{Y})}{\phi'(R_0)}\right) > 0,\end{aligned}$$

where the first inequality used $(1-t_0) > (1-t_1)$, and the second follows because $t_0 < 1$, $V_0 > \bar{Y}$, and ϕ is strictly increasing. We conclude that $\pi(v_0; t_1)$ is a strictly increasing function of t_1, which implies that $\pi(v_0) \equiv \pi(v_0; t_1) > \pi(v_0; t_0) = 0$ for all $v_0 > \bar{y}$. For $v_0 = \bar{y}$, it is clear by inspection that $\pi(\bar{y}; t_1) \in (0, \tau)$. Q.E.D.

Proof of Proposition 2. Equations (VE) and (2) imply (3). Because $q_0(\bar{p})$ is known, the assumed value of η implies a unique value of ϵ given by (3), so that ϵ is point identified. If no value for η is assumed, then it is still known that $\eta \in [-1, 0]$ due to the Engel aggregation condition, as noted in Section 2. Substituting the extremal points $\eta = 0$ and $\eta = -1$ gives the bounds on ϵ shown in (4).

To see that the bounds are sharp, fix any ϵ within the bounds and let $\eta \in [-1, 0]$ be such that (3) holds. Then

$$\nu(y_0, \epsilon, \eta) = y_0 - (q_0(\bar{p}) - \bar{y}) + \eta\,(\pi(q_0(\bar{p})) - \pi(y_0))$$

which is a strictly increasing function of y_0. Consider a distribution of y_0 that is uniform on $[\bar{y}, q_0(\bar{p})]$ and coincides with a transformed version of G_1 above $q_0(\bar{p})$, namely:

$$G_0(z) = \mathbb{1}[\bar{y} \le z < q_0(\bar{p})] \left(\frac{z - \bar{y}}{q_0(\bar{p}) - \bar{y}} \right) p + \mathbb{1}[z \ge q_0(\bar{p})] G_1\left(\nu(z, \epsilon, \eta)\right).$$

This distribution is consistent with the bunching quantile because

$$G_0(q_0(\bar{p})) = G_1(\nu(q_0(\bar{p}), \epsilon, \eta)) = G_1(\bar{y}) = \bar{p}. \tag{20}$$

It also reproduces the distribution of y_1 because for any $y \ge \bar{y}$,

$$\mathbb{P}[y_1 \le y] = \mathbb{P}[\nu(y_0, \epsilon, \eta) \le y] = \mathbb{P}[y_0 \le \nu^{-1}(y, \epsilon, \eta)] = G_0(\nu^{-1}(y, \epsilon, \eta)) = G_1(y).$$

This shows that the bounds are sharp because for every value of ϵ in (4) there is a distribution of y_0 that is consistent with both G_1 and $q_0(\bar{p})$. Q.E.D.

Proof of Proposition 3. From (6) evaluated at $p = p'$ and $p = p''$ we get two equations in two unknowns:

$$\begin{aligned} &\tau\epsilon + \eta\pi(q_0(p')) = q_0(p') - q_1(p') \\ \text{and} \quad &\tau\epsilon + \eta\pi(q_0(p'')) = q_0(p'') - q_1(p''). \end{aligned}$$

This system of equations has a solution because both π and q_0 are strictly increasing functions, the former by definition in Proposition 1 and the latter because y_0 is assumed to be continuously distributed. Solving the system yields (7)–(8). Q.E.D.

C Appendix figures and tables

Figure C.1: Comparison of self-employed workers and the general population

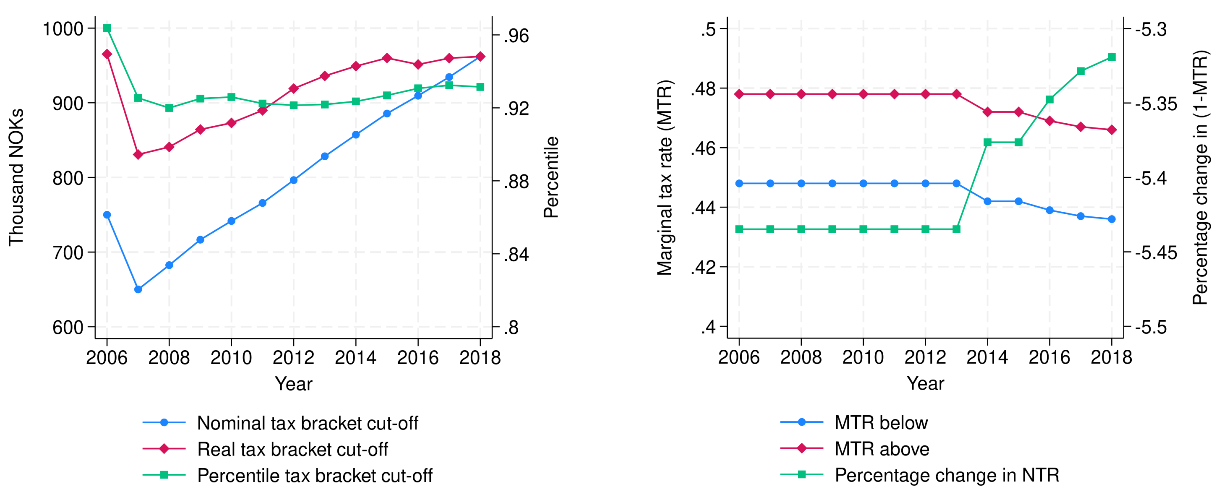
	Population	Self-employed
Age	42.8	45.1
Female	0.49	0.31
College	0.37	0.29
Taxable income	492,352	552,684
Wage income	426,194	42,144
Capital income	25,478	36,595
Observations	31,880,219	1,540,136

NOK is Norwegian Krone, which exchanged for between 5 and 9 per USD over the sample period. Population is the entire Norwegian prime-working-age population between 2006 and 2018. An individual is self-employed in a given year if they have business income larger than wage income that year.

Figure C.2: Top income tax over time

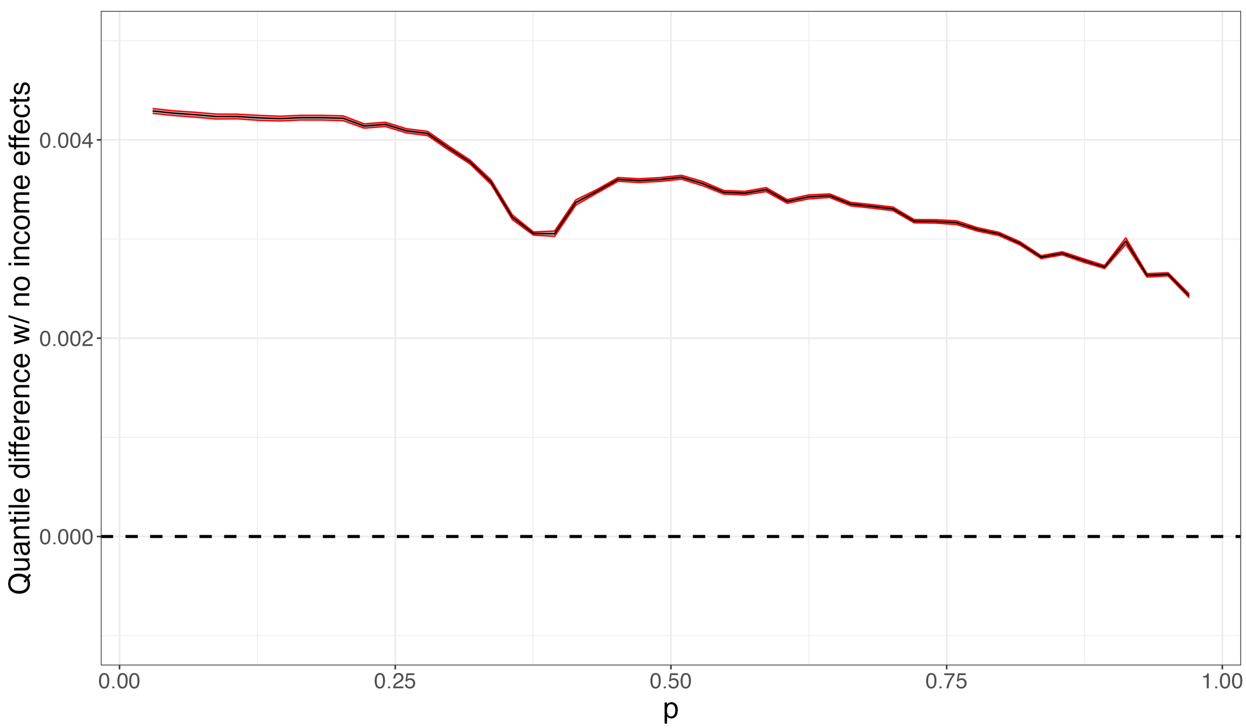
(a) Top tax bracket cutoff

(b) Marginal tax rates



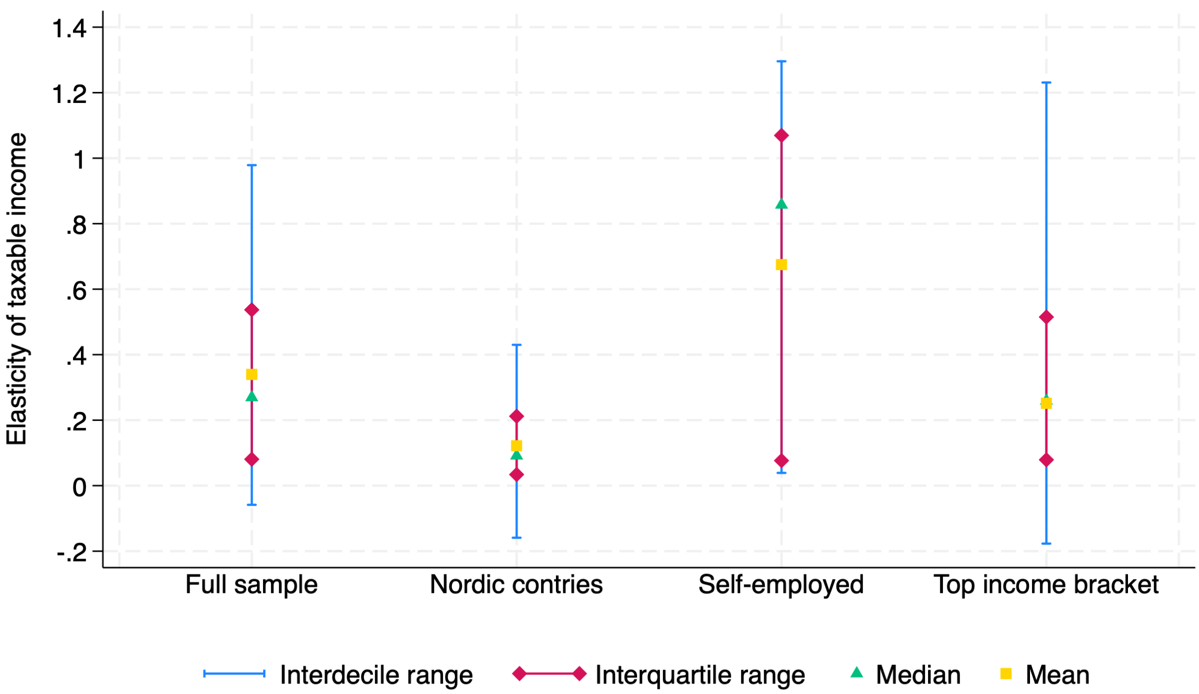
Notes: The graphs plot the evolution of the Norwegian top tax bracket over time. Panel (a) shows how the cutoff value varies over time in nominal and real terms, as well as a percentile in the earnings distribution among the self-employed. Panel (b) plots the marginal tax rates in the two highest tax brackets over time.

Figure C.3: Estimates of the sharp testable implication in Proposition 4



Notes: The figure plots $\zeta(q_0(p);\eta) - q_1(p)$ as a function of p with $\eta = 0$ in black, with pointwise 95% confidence intervals in red. Proposition 4 says the model is misspecified if and only if this function is smaller than zero for some p. The estimated function is always above zero, so any test will fail to reject this premise.

Figure C.4: Elasticities from Neisser (2021)



Notes: This figure plots interdecile- and quartile ranges, median, and mean elasticities of taxable income estimates for four different subsamples using data from Neisser (2021). "Full sample" refers to all 1,720 estimates included in the meta-analysis. Nordic countries include estimates from Denmark, Finland, Norway, and Sweden (N = 292). Self-employed contains estimates from samples of self-employed (N = 20), and top income bracket refers to studies where a top-income tax rate is introduced (N = 431).

Table C.1: Using different amounts of data

		Data until 5th quantile ($\bar{\bar{p}}_0 = 0.05$)				Data until 20th quantile ($\bar{\bar{p}}_0 = 0.2$)			
		Compensated elasticity		Uncompensated elasticity		Compensated elasticity		Uncompensated elasticity	
η_{MIN}	ϵ_{MAX}	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
0	3	$[0.000, 0.635]$	$(0.000, 0.637)$			$[0.000, 0.217]$	$(0.000, 0.218)$		
0	1.5	$[0.000, 0.349]$	$(0.000, 0.350)$			$[0.000, 0.145]$	$(0.000, 0.146)$		
0	1	$[0.000, 0.254]$	$(0.000, 0.255)$			$[0.000, 0.121]$	$(0.000, 0.122)$		
-1	3	$[0.000, 0.667]$	$(0.000, 0.670)$	$[-1.000, 0.635]$	$(-1.000, 0.637)$	$[0.000, 0.287]$	$(0.000, 0.288)$	$[-1.000, 0.217]$	$(-1.000, 0.218)$
-1	1.5	$[0.000, 0.381]$	$(0.000, 0.383)$	$[-1.000, 0.349]$	$(-1.000, 0.350)$	$[0.000, 0.216]$	$(0.000, 0.216)$	$[-1.000, 0.145]$	$(-1.000, 0.146)$
-1	1	$[0.000, 0.286]$	$(0.000, 0.287)$	$[-1.000, 0.254]$	$(-1.000, 0.255)$	$[0.000, 0.192]$	$(0.000, 0.192)$	$[-1.000, 0.121]$	$(-1.000, 0.122)$
-0.5	3	$[0.000, 0.651]$	$(0.000, 0.653)$	$[-0.500, 0.635]$	$(-0.500, 0.637)$	$[0.000, 0.252]$	$(0.000, 0.253)$	$[-0.500, 0.217]$	$(-0.500, 0.218)$
-0.5	1.5	$[0.000, 0.365]$	$(0.000, 0.367)$	$[-0.500, 0.349]$	$(-0.500, 0.350)$	$[0.000, 0.181]$	$(0.000, 0.181)$	$[-0.500, 0.145]$	$(-0.500, 0.146)$
-0.5	1	$[0.000, 0.270]$	$(0.000, 0.271)$	$[-0.500, 0.254]$	$(-0.500, 0.255)$	$[0.000, 0.157]$	$(0.000, 0.157)$	$[-0.500, 0.121]$	$(-0.500, 0.122)$
-0.3	3	$[0.000, 0.644]$	$(0.000, 0.647)$	$[-0.300, 0.635]$	$(-0.300, 0.637)$	$[0.000, 0.238]$	$(0.000, 0.239)$	$[-0.300, 0.217]$	$(-0.300, 0.218)$
-0.3	1.5	$[0.000, 0.359]$	$(0.000, 0.360)$	$[-0.300, 0.349]$	$(-0.300, 0.350)$	$[0.000, 0.166]$	$(0.000, 0.167)$	$[-0.300, 0.145]$	$(-0.300, 0.146)$
-0.3	1	$[0.000, 0.264]$	$(0.000, 0.265)$	$[-0.300, 0.254]$	$(-0.300, 0.255)$	$[0.000, 0.143]$	$(0.000, 0.143)$	$[-0.300, 0.121]$	$(-0.300, 0.122)$
-0.1	3	$[0.000, 0.638]$	$(0.000, 0.640)$	$[-0.100, 0.635]$	$(-0.100, 0.637)$	$[0.000, 0.224]$	$(0.000, 0.225)$	$[-0.100, 0.217]$	$(-0.100, 0.218)$
-0.1	1.5	$[0.000, 0.352]$	$(0.000, 0.354)$	$[-0.100, 0.349]$	$(-0.100, 0.350)$	$[0.000, 0.152]$	$(0.000, 0.153)$	$[-0.100, 0.145]$	$(-0.100, 0.146)$
-0.1	1	$[0.000, 0.257]$	$(0.000, 0.258)$	$[-0.100, 0.254]$	$(-0.100, 0.255)$	$[0.000, 0.129]$	$(0.000, 0.129)$	$[-0.100, 0.121]$	$(-0.100, 0.122)$

Notes: This table reports bounds on average compensated and uncompensated elasticities for top-earners with y_0 below the 5th (first set of columns) and 20th (second set of columns) quantiles when using knowledge of the distribution of y_0 only up until that quantile and using full knowledge of the distribution of y_1. We consider bounds under different choices of ϵ_{MAX} and η_{MIN}, where $\eta_{\text{MIN}} = 0$ denotes the assumption of no income effects. Confidence intervals are constructed using the method of Imbens and Manski (2004) with the bootstrap to account for estimating the pre-kink distribution.

Supplemental Appendix

SA.1 Welfare derivations

In Section 3, we transformed our compensated elasticity estimates with homogeneous elasticities into the implied revenue-maximizing tax rate and excess burden of taxation. This section provides the derivations underlying these figures.

SA.1.1 Revenue-maximizing top-income tax rate

We first express the top-income tax revenue as a function of the top-income tax rate t. Proposition 1 provided an expression for the virtual earnings choice under a linear tax function that intersects tax system 1 at $\bar{Y}$. Write that solution as an explicit function of the marginal tax rate t:

$$\nu(v_0, \epsilon, \eta; t) \equiv v_0 - \epsilon\tau(t) - \eta\pi(v_0; t),$$

where

$$\tau(t) \equiv \log\left(\frac{1-t_0}{1-t}\right) \quad \text{and} \quad \pi(v_0; t) \equiv \tau(t) - \frac{\phi(R_0 + \bar{Y}(t - t_0)) - \phi(R_0)}{(1-t_0)\exp(v_0)\phi'(R_0)}.$$

The tax revenue collected from the top-income tax bracket is

$$\textsc{tr}(t; \epsilon, \eta) \equiv t\,\mathbb{E}\left[\max\left\{0, \exp(\nu(v_0, \epsilon, \eta; t)) - \bar{Y}\right\}\right],$$

where ϵ and η are deterministic constants. The distribution of v_0 is given by G_0 for the top bracket case. In the homogeneous case, identifying ϵ, η, and $G_0(z)$ for $z \in [\bar{y}, q_0(\bar{p})]$ is sufficient to pin down $G_0(z)$ for all z:

$$G_0(z) = G_1\left(\nu(z, \epsilon, \eta, t_1)\right) \quad \text{for all } z > q_0(\bar{p}).$$

In this case, $\textsc{tr}(t; \epsilon, \eta)$ is also identified, as is the value of t that maximizes revenue:

$$t^*(\epsilon, \eta) \equiv \underset{t \in [t_0, 1]}{\arg\max}\ \textsc{tr}(t; \epsilon, \eta).$$

An estimate of the revenue-maximizing tax rate can be obtained by solving this problem with estimates of ϵ and η.

SA.1.2 Excess burden of taxation

Defining the excess burden of taxation requires introducing the expenditure function:

$$E(t, \bar{u}) \equiv \min_{C,Y} \; C - (1-t)Y \text{ subject to } U(C, Y) \geq \bar{u}. \tag{21}$$

The solution to this problem is the compensated earnings function, which we denote by $Y^c(t, \bar{u})$. Deadweight loss is then defined as the net-of-revenue expenditure needed to ensure that an individual can obtain utility $\bar{u}$ under tax rate t:

$$\text{DWL}(t, \bar{u}) \equiv E(t, \bar{u}) - tY^c(t, \bar{u}).$$

The marginal deadweight loss is the derivative of DWL with respect to t:

$$\begin{aligned}
\text{MDWL}(t, \bar{u}) &= \frac{\partial}{\partial t} E(t, \bar{u}) - t\frac{\partial}{\partial t} Y^c(t, \bar{u}) - Y^c(t, \bar{u}) \\
&= -t\frac{\partial}{\partial t} Y^c(t, \bar{u}) \\
&= t\frac{\partial}{\partial (1-t)} Y^c(t, \bar{u}) = \frac{tY^c(t, \bar{u})}{1-t} \left(\frac{\partial Y^c(t, \bar{u})}{\partial (1-t)} \frac{1-t}{Y^c(t, \bar{u})} \right),
\end{aligned}$$

where the second equality follows from the envelope theorem (Shephard's Lemma).

When $\bar{u} = U^\star(1-t_0, R_0)$ is set to the utility obtained under tax system 0,

$$\text{MDWL}(t_0, U^\star(1-t_0, R_0)) = \frac{t_0 Y^\star(1-t_0, R_0)\epsilon(1-t_0, R_0)}{1-t_0} \equiv \frac{t_0 V_0 \epsilon}{1-t_0}.$$

It is common to normalize marginal deadweight loss by marginal tax revenue:

$$\text{MTR}(t, \bar{u}) \equiv \frac{\partial}{\partial t}\left(tY^c(t, \bar{u}) \right) = Y^c(t, \bar{u}) + t\frac{\partial}{\partial t} Y^c(t, \bar{u}) = Y^c(t, \bar{u}) - \text{MDWL}(t, \bar{u}).$$

The excess burden is the ratio of MDWL to MTR, which measures government's cost of increasing the marginal tax rate relative to the additional revenue raised, holding the consumer indifferent. Under tax system 0 this is

$$\begin{aligned}
\text{EB}(t_0, U^\star(1-t_0, R_0)) &\equiv \frac{\text{MDWL}(t_0, U^\star(1-t_0, R_0))}{\text{MTR}(t_0, U^\star(1-t_0, R_0))} \\
&= \left(\frac{t_0 \epsilon V_0}{1-t_0} \right) \left(V_0 - \frac{t_0 \epsilon V_0}{1-t_0} \right)^{-1} = \frac{t_0 \epsilon}{1 - t_0(1+\epsilon)}.
\end{aligned}$$

SA.2 Proofs for Sections 5–6

Our proofs make repeated use of the following three lemmas.

Lemma EX. (Existence of a distribution of compensated elasticity) Suppose that $\eta \in [-1, 0]$ is deterministic. Let $\mathcal{F}^{\dagger}_{\mathrm{D}}$ denote the specification of $\mathcal{F}^{\dagger}$ given by Condition D with this choice of η. Let $\mathcal{F}^{\star}_{\mathrm{D}}$ denote $\mathcal{F}^{\star}$ with $\mathcal{F}^{\dagger} = \mathcal{F}^{\dagger}_{\mathrm{D}}$. Let Φ_0 be a strictly increasing and absolutely continuous distribution such that $\Phi_0(y) = G_0(y)$ for all $y \in (\bar{y}, \bar{\bar{y}}_0)$. Let Φ_1 be a strictly increasing distribution that is absolutely continuous on $(\bar{y}, \bar{\bar{y}}_1)$, satisfies $\Phi_1(y) = G_1(y)$ for all $y \in [\bar{y}, \bar{\bar{y}}_1)$, and has $\lim_{y \uparrow \bar{y}} \Phi_1(y) = 0$, so that it places no mass below $\bar{y}$. Suppose that $\Phi_0(\zeta^{-1}(y; \eta)) \le \Phi_1(y)$ for all $y \ge \bar{y}$.

Let v_0 be a random variable with distribution Φ_0. Let H_0 denote the distribution of $\zeta(v_0; \eta)$. Given any $e \ge 0$, define the random variable

$$\epsilon = \frac{1}{\tau}\left[\zeta(v_0;\eta) - \Phi_1^{-1}(H_0(\zeta(v_0;\eta)))\right] + \frac{e}{\tau}\mathbb{1}[H_0(\zeta(v_0;\eta)) \le \bar{p}], \tag{EX-ϵ}$$

where Φ_1^{-1} is the generalized inverse of Φ_1 and $\bar{p} \equiv \Phi_1(\bar{y}) = G_1(\bar{y})$. Let F denote the joint distribution of (ϵ, η, v_0) with η deterministic. Then $F \in \mathcal{F}^{\star}_{\mathrm{D}}$. Moreover,

$$\mathbb{E}_F[\epsilon | B = b] = \begin{cases} \frac{1}{\tau}(\mathbb{E}[\zeta(v_0;\eta)|v_0 > q_0(\bar{p})] - \mathbb{E}[v_1|v_1 > \bar{y}]) & \text{if } b = 0, \\ \frac{1}{\tau}(\mathbb{E}[\zeta(v_0;\eta)|v_0 \le q_0(\bar{p})] - \bar{y} + e) & \text{if } b = 1, \end{cases} \tag{EX-ϵ-$\mathbb{E}$}$$

where v_1 is a random variable with distribution Φ_1 and $B \equiv \mathbb{1}[\nu(v_0, \epsilon, \eta) \le \bar{y}]$. The expressions in (EX-ϵ-$\mathbb{E}$) are functions of Φ_0 and Φ_1.

Proof of Lemma EX. We simplify the notation by defining $z_0 \equiv \zeta(v_0; \eta)$ and $u_0 \equiv H_0(z_0)$. Then (EX-ϵ) becomes

$$\epsilon \equiv \frac{1}{\tau}\left(z_0 - \Phi_1^{-1}(u_0)\right) + \frac{e}{\tau}\mathbb{1}[u_0 \le \bar{p}]. \tag{22}$$

The distribution of u_0 is uniform on $[0, 1]$, because z_0 is continuously distributed with distribution function H_0. Notice that H_0 satisfies

$$H_0(y) = \mathbb{P}[\zeta(v_0;\eta) \le y] = \mathbb{P}[v_0 \le \zeta^{-1}(y;\eta)] = \Phi_0(\zeta^{-1}(y;\eta)) \le \Phi_1(y),$$

which is an inequality that holds for all $y \ge \bar{y}$, by assumption. This implies that

$$\mathbb{P}[z_0 \ge \Phi_1^{-1}(u_0)] = \mathbb{P}[\Phi_1(z_0) \ge H_0(z_0)] = 1,$$

which implies that $\epsilon \ge 0$ with probability one. We conclude that $F \in \mathcal{F}^{\dagger}_{\mathrm{D}}$.

Rearranging (VE) and substituting the definition of ϵ gives virtual earnings

$$\nu(v_0, \epsilon, \eta) = v_0 - \eta\pi(v_0) - \epsilon\tau = z_0 - \epsilon\tau = \Phi_1^{-1}(u_0) - e\mathbb{1}[u_0 \le \bar{p}]. \tag{23}$$

Because $\Phi_1^{-1}(u_0) \geq \bar{y} = q_1(\bar{p})$ and $e \geq 0$, we have

$$\nu(v_0, \epsilon, \eta) \leq \bar{y} \quad \Leftrightarrow \quad u_0 \leq \bar{p} \quad \Leftrightarrow \quad v_0 \leq \Phi_0^{-1}(\bar{p}), \tag{24}$$

where the second equivalence follows because u_0 is uniformly distributed on $[0, 1]$, $u_0 \equiv H_0(\zeta(v_0; \eta))$ is a strictly increasing transformation of v_0 (Proposition 1), and v_0 has distribution Φ_0 with quantiles Φ_0^{-1}. Using (23) and the first two equivalences of (24), we conclude that

$$\begin{aligned} \mathbb{P}_F[\nu(v_0, \epsilon, \eta) \leq y] &= \mathbb{P}_F[\bar{y} < \nu(v_0, \epsilon, \eta) \leq y] + \mathbb{P}_F[\nu(v_0, \epsilon, \eta) \leq \bar{y}] \\ &= \mathbb{P}_F[\Phi_1^{-1}(u_0) \leq y] - \mathbb{P}_F[\Phi_1^{-1}(u_0) \leq \bar{y}] + \mathbb{P}_F[u_0 \leq \bar{p}] \\ &= \Phi_1(y) - \Phi_1(\bar{y}) + \bar{p} = \Phi_1(y), \end{aligned} \tag{25}$$

where the third equality follows from the definition of a quantile as a generalized inverse of a right-continuous function (e.g. Embrechts and Hofert, 2013, Proposition 1(5)), together with u_0 being uniformly distributed on $[0, 1]$. From (25), and the assumption that $\Phi_1(y) = G_1(y)$ for all $y \in [\bar{y}, \bar{\bar{y}}_1)$, we conclude that $\mathbb{P}_F[\nu(v_0, \epsilon, \eta) \leq y] = G_1(y)$ for all $y \in [\bar{y}, \bar{\bar{y}}_1)$. We have also constructed F such that $\mathbb{P}_F[v_0 \leq y] = \Phi_0(y)$, which is equal to $G_0(y)$ for all $y \in (\bar{y}, \bar{\bar{y}}_0)$, by assumption. We conclude that $F \in \mathcal{F}_{\text{D}}^{\star}$.

We show (EX-ϵ-$\mathbb{E}$) by directly computing the conditional expectation of (22), noting that the event $B = 1$ is equivalent to (24):

$$\mathbb{E}_F[\epsilon | B = b] = \begin{cases} \frac{1}{\tau}(\mathbb{E}[\zeta(v_0; \eta) | v_0 > q_0(\bar{p})] - \mathbb{E}[\Phi_1^{-1}(u_0) | u_0 > \bar{p}]) & \text{if } b = 0, \\ \frac{1}{\tau}(\mathbb{E}[\zeta(v_0; \eta) | v_0 \leq q_0(\bar{p})] - \mathbb{E}[\Phi_1^{-1}(u_0) | u_0 \leq \bar{p}] + e) & \text{if } b = 1. \end{cases} \tag{26}$$

By assumption, $\Phi_1(\bar{y}) = G_1(\bar{y}) \equiv \bar{p}$, so $\Phi_1^{-1}(u) = \bar{y}$ for all $u \leq \bar{p}$. This implies that $\mathbb{E}[\Phi_1^{-1}(u_0) | u_0 \leq \bar{p}] = \bar{y}$, which yields the expression for $b = 1$. For $b = 0$, a change of variables implies that

$$\mathbb{E}[\Phi_1^{-1}(u_0) | u_0 > \bar{p}] = \int_{\bar{p}}^{1} \Phi_1^{-1}(u) \frac{du}{1 - \bar{p}} = \int_{\bar{y}}^{\infty} y \frac{\Phi_1'(y) dy}{1 - \Phi_1(\bar{y})} = \mathbb{E}[v_1 | v_1 > \bar{y}].$$

Q.E.D.

Lemma HM. (Horowitz-Manski bounds) Suppose that B is a binary random variable with $\mathbb{P}[B = 1] = \bar{p}$ known. Suppose that η is deterministic and let $z_0 \equiv \zeta(v_0; \eta)$. Suppose that v_0 has a strictly increasing and absolutely continuous distribution function Φ_0 with quantile function Φ_0^{-1}. Then

$$\mathbb{E}[z_0|v_0 \leq \Phi_0^{-1}(1-\bar{p})] \leq \mathbb{E}[z_0|B=0] \leq \mathbb{E}[z_0|v_0 > \Phi_0^{-1}(\bar{p})]$$
$$\text{and} \quad \mathbb{E}[z_0|v_0 \leq \Phi_0^{-1}(\bar{p})] \leq \mathbb{E}[z_0|B=1] \leq \mathbb{E}[z_0|v_0 > \Phi_0^{-1}(1-\bar{p})].$$

Proof of Lemma HM. This follows directly from Proposition 4 of Horowitz and Manski (1995) after noting from Proposition 1 that $\zeta(v_0;\eta)$ is a strictly increasing function of v_0 for any fixed η. Because v_0 is continuously distributed with distribution Φ_0, this ensures that z_0 is smaller than its $\bar{p}$th quantile if and only if v_0 is smaller than $\Phi_0^{-1}(\bar{p})$, and similarly for the other events. Q.E.D.

Lemma IDT. (Identified set under Condition T) Maintain Condition T. Suppose that $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables that are distributed like F. Let $v_1 \equiv \nu(v_0, \epsilon, \eta)$. Then $\mathbb{P}[v_0 \leq y] = G_0(y)$ for all $y \geq \bar{y}$ and $\mathbb{P}[v_1 \leq y] = G_1(y)$ for all $y \geq \bar{y}$. The distribution of y_1 first-order stochastically dominates the distribution of v_1.

Proof of Lemma IDT. The equalities of the distribution functions are direct implications of the definition of $\mathcal{F}^\star$ under Condition T. That the distribution of y_1 first-order stochastically dominates that of v_1 then follows because $\mathbb{P}[y_1 < \bar{y}] = 0$. Q.E.D.

Proof of Proposition 4. Suppose that the model is not misspecified, so that $\mathcal{F}^\star$ is non-empty. Let $F \in \mathcal{F}^\star$. Then for all y,

$$\begin{aligned} G_1(y) &= \mathbb{P}_F[\nu(v_0, \epsilon, \eta) \leq y] \\ &= \mathbb{P}_F[\zeta(v_0;\eta) - \tau\epsilon \leq y] \\ &\geq \mathbb{P}_F[\zeta(v_0;\eta) \leq y] = \mathbb{P}[\zeta(y_0;\eta) \leq y], \end{aligned} \tag{27}$$

where the first equality used the definition of $\mathcal{F}^\star$ under Condition T, the second used (VE), the inequality follows because $F \in \mathcal{F}^\dagger$ have $\epsilon \geq 0$ with probability one, and the final equality again used $F \in \mathcal{F}^\star$ under Condition T.

For the opposite implication, suppose that $\mathbb{P}[\zeta(y_0;\eta) \leq y] \leq G_1(y)$ for all y. Given Condition D, this can be rewritten as $\mathbb{P}[\zeta(y_0;\eta) \leq y] = G_0(\zeta^{-1}(y;\eta)) \leq G_1(y)$ for all y. The existence of an $F \in \mathcal{F}^\star$ now follows from Lemma EX by taking $\Phi_0 = G_0$ and $\Phi_1 = G_1$. Q.E.D.

Proof of Proposition 5. We first show that the bounds are valid. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Define

$$v_1 \equiv \nu(v_0, \epsilon, \eta) \equiv v_0 - \epsilon\tau - \eta\pi(v_0) \equiv z_0 - \epsilon\tau, \tag{VE}$$

where $z_0 \equiv v_0 - \eta\pi(v_0) \equiv \zeta(v_0; \eta)$.

Lemma IDT shows that $\mathbb{E}[v_1] \leq \mathbb{E}[y_1]$, which upon rearranging (VE) gives:

$$\mathbb{E}[\epsilon] = \frac{1}{\tau}\left(\mathbb{E}[z_0] - \mathbb{E}[v_1]\right) \geq \frac{1}{\tau}\left(\mathbb{E}[z_0] - \mathbb{E}[y_1]\right).$$

Substituting the definition of z_0 produces (P5-LB). Next, we apply Lemma HM to obtain

$$\begin{aligned}
&\mathbb{E}[\epsilon|B = 0] \\
&\quad= \frac{1}{\tau}\left(\mathbb{E}[z_0|B = 0] - \mathbb{E}[v_1|B = 0]\right) \\
&\quad\leq \frac{1}{\tau}\left(\mathbb{E}[z_0|v_0 > q_0(\bar{p})] - \mathbb{E}[v_1|v_1 > \bar{y}]\right) = \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0; \eta)|y_0 > q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right),
\end{aligned}$$

where the final equality used $F \in \mathcal{F}^\star$ with Lemma IDT. This establishes (P5-UB-B0). The lower bound (P5-LB-B1) also follows from Lemma HM after noting that $\mathbb{E}[v_1|B = 1] \leq \bar{y}$:

$$\mathbb{E}[\epsilon|B = 1] = \frac{1}{\tau}\left(\mathbb{E}[z_0|B = 1] - \mathbb{E}[v_1|B = 1]\right) \geq \frac{1}{\tau}\left(\mathbb{E}[z_0|v_0 \leq q_0(\bar{p})] - \bar{y}\right).$$

Substituting the definition of z_0 produces (P5-LB-B1), because $F \in \mathcal{F}^\star$.

To establish (P5-LB-B0), let $\Phi_0 \equiv \mathbb{P}[v_0 \leq v]$ and $\Phi_{0|b}(v) \equiv \mathbb{P}[v_0 \leq v|B = b]$ for $b = 0, 1$, then use the mixture decomposition (10) with Φ_0 in place of G_0:

$$\Phi_{0|0}(v) = \frac{\Phi_0(v) - \Phi_{0|1}(v)\bar{p}}{1 - \bar{p}} \leq \min\left\{1, \frac{\Phi_0(v)}{1 - \bar{p}}\right\} = \min\left\{1, \frac{G_0(v)}{1 - \bar{p}}\right\}, \tag{28}$$

where the final equality follows from $F \in \mathcal{F}^\star$. In addition to this inequality, we also have from (VE) that

$$\begin{aligned}
\Phi_{0|0}(v) &= \mathbb{P}[\zeta(v_0; \eta) \leq \zeta(v; \eta)|B = 0] \\
&= \mathbb{P}[v_1 + \epsilon\tau \leq \zeta(v; \eta)|B = 0] \\
&\leq \mathbb{P}[v_1 \leq \zeta(v; \eta)|B = 0] = G_{1|0}(\zeta(v; \eta)),
\end{aligned} \tag{29}$$

where the inequality follows because $\mathbb{P}[\epsilon \geq 0] = 1$ and the equality is because $F \in \mathcal{F}^\star$.

Together, (28) and (29) imply that

$$\Phi_{0|0}(v) \leq \min\{G_0(v)/(1-\bar{p}), G_{1|0}(\zeta(v;\eta))\} \equiv \bar{G}(v).$$

We obtain (P5-LB-B0) from this bound because $\zeta(\cdot;\eta)$ is strictly increasing:

$$\begin{aligned}
&\mathbb{E}[\epsilon|B=0] \\
&\quad = \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;\eta)|B=0] - \mathbb{E}[v_1|B=0]\right) \\
&\quad = \frac{1}{\tau}\left(\int \zeta(y;\eta)\, d\Phi_{0|0}(y) - \mathbb{E}[v_1|v_1>\bar{y}]\right) \geq \frac{1}{\tau}\left(\int \zeta(y;\eta)\, d\bar{G}(y) - \mathbb{E}[y_1|y_1>\bar{y}]\right),
\end{aligned}$$

where the final equality used $F \in \mathcal{F}^\star$.

We now establish that the bounds are sharp by applying Lemma EX. We have assumed that the model is not misspecified, which Proposition 4 shows is equivalent to assuming that $\mathbb{P}[\zeta(y_0;\eta) \leq y] = G_0(\zeta^{-1}(y;\eta)) \leq G_1(y)$ for all $y \geq \bar{y}$. The assumptions of Lemma EX are therefore satisfied with $\Phi_0 = G_0$ and $\Phi_1 = G_1$.

For any $e \geq 0$, Lemma EX shows that there exists an $F \in \mathcal{F}^\star_{\mathrm{D}} = \mathcal{F}^\star$ such that (EX-ϵ-$\mathbb{E}$) holds. Taking $e = 0$, the expression for $b = 0$ in (EX-ϵ-$\mathbb{E}$) becomes the upper bound (P5-UB-B0), while the expression for $b = 1$ in (EX-ϵ-$\mathbb{E}$) becomes the lower bound (P5-LB-B1). The lower bound (P5-LB) is also achieved when $e = 0$ by applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$).

On the other hand, taking e to be arbitrarily large in (EX-ϵ-$\mathbb{E}$) makes $\mathbb{E}_F[\epsilon|B=1]$ and therefore $\mathbb{E}_F[\epsilon]$ arbitrarily large. This implies that the sharp upper bounds on $\mathbb{E}[\epsilon|B=1]$ and $\mathbb{E}[\epsilon]$ are infinite. The sharpness of the lower bound on $\mathbb{E}[\epsilon|B=0]$ under the stated single-crossing condition follows as a special case of the more general result proven in Appendix SA.4. Q.E.D.

Proof of Proposition 6. Suppose that $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be distributed like F. The law of iterated expectations gives

$$\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon|B=1]\bar{p} + \mathbb{E}[\epsilon|B=0](1-\bar{p}). \tag{30}$$

The claimed upper bound follows after bounding $\mathbb{E}[\epsilon|B=1]$ by ϵ_{MAX} and $\mathbb{E}[\epsilon|B=0]$ by (P5-UB-B0). Note that (P5-UB-B0) is still a valid bound because Condition DB restricts $\mathcal{F}^\dagger$ to be a strictly smaller set than it is under Condition D.

To establish sharpness, assume that the model is not misspecified. Because Condition DB is more restrictive than D, Proposition 4 implies that $\mathbb{P}[\zeta(y_0;\eta) \leq y] = G_0(\zeta^{-1}(y;\eta)) \leq G_1(y)$ for all $y \geq \bar{y}$. The assumptions of Lemma EX are therefore satisfied with $\Phi_0 = G_0$ and $\Phi_1 = G_1$. We conclude that for any $e \geq 0$, there exists an $F \in \mathcal{F}_{\mathrm{D}}^{\star}$ with $\mathbb{E}_F[\epsilon|B = b]$ given by (EX-ϵ-$\mathbb{E}$).

Taking $e = 0$ produces an $F_{\mathrm{LB}} \in \mathcal{F}_{\mathrm{D}}^{\star}$ for which $\mathbb{E}_{F_{\mathrm{LB}}}[\epsilon|B = 1]$ obtains the lower bound (P5-LB-B1), which is sharp when $\mathcal{F}^{\dagger} = \mathcal{F}_{\mathrm{D}}^{\dagger}$. Because $\mathcal{F}^{\star} \subseteq \mathcal{F}_{\mathrm{D}}^{\star}$ and $\mathcal{F}^{\star}$ is non-empty by hypothesis, we conclude that

$$\epsilon_{\mathrm{MAX}} \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\eta)|y_0 \leq q_0(\bar{p})] - \bar{y}\right) = \mathbb{E}_{F_{\mathrm{LB}}}[\epsilon|B = 1]. \tag{31}$$

The lower bound (P5-LB) is obtained by taking $e = 0$ and applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$). To obtain the upper bound, set

$$\begin{aligned} e^{\star} &\equiv \tau\epsilon_{\mathrm{MAX}} + \bar{y} - \mathbb{E}[z_0|y_0 \leq q_0(\bar{p})] \\ &\geq \tau\left(\frac{1}{\tau}\left(\mathbb{E}[z_0|y_0 \leq q_0(\bar{p})] - \bar{y}\right)\right) + \bar{y} - \mathbb{E}[z_0|y_0 \leq q_0(\bar{p})] = 0, \end{aligned}$$

where $z_0 \equiv \zeta(y_0;\eta)$ and the inequality uses (31). Applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$) produces

$$\begin{aligned} \mathbb{E}_F[\epsilon] &= \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;\eta)|v_0 > q_0(\bar{p})] - \mathbb{E}[v_1|v_1 > \bar{y}]\right)(1-\bar{p}) \\ &\quad + \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;\eta)|v_0 \leq q_0(\bar{p})] - \bar{y} + e^{\star}\right)\bar{p} \\ &= (1-\bar{p})\frac{1}{\tau}\left(\mathbb{E}[z_0|y_0 > q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right) + \bar{p}\frac{1}{\tau}\left(\mathbb{E}[z_0|y_0 \leq q_0(\bar{p})] - \bar{y} + e^{\star}\right) \\ &= \epsilon_{\mathrm{MAX}}\bar{p} + \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0;\eta)|y_0 > q_0(\bar{p})\right] - \mathbb{E}[y_1|y_1 > \bar{y}]\right)(1-\bar{p}), \end{aligned}$$

where the second equality follows because we took $\Phi_0 = G_0$ and $\Phi_1 = G_1$ in Lemma EX. Moreover, (EX-ϵ-$\mathbb{E}$) and the definition of $e^{\star}$ imply that $\mathbb{E}_F[\epsilon|B = 1] = \epsilon_{\mathrm{MAX}}$. We conclude that $F \in \mathcal{F}^{\star}$, which implies that the bound is sharp. Q.E.D.

Proof of Proposition 7. Suppose that the model is not misspecified, so that $\mathcal{F}^{\star}$ is non-empty. Let $F \in \mathcal{F}^{\star}$. In (27) from the proof of Proposition 4, we showed that $G_1(y) \geq \mathbb{P}_F[\zeta(v_0;\eta) \leq y]$ for all y. This inequality only used $\epsilon \geq 0$, so it continues to hold when $\mathcal{F}^{\dagger} = \mathcal{F}$ allows for η to be stochastic. We combine this inequality with the observation that $\zeta(v_0;\eta) \equiv v_0 - \eta\pi(v_0)$ is a decreasing function of $\eta \in [-1, 0]$ for any v_0, which is a consequence of π being non-negative (Proposition 1). Then

$$G_1(y) \geq \mathbb{P}_F[\zeta(v_0;\eta) \leq y] \geq \mathbb{P}_F[\zeta(v_0;-1) \leq y] = \mathbb{P}[\zeta(y_0;-1) \leq y],$$

where the final equality uses $F \in \mathcal{F}^{\star}$ under Condition T.

For the opposite implication, suppose that $\mathbb{P}[\zeta(y_0;-1) \leq y] \leq G_1(y)$ for all y. This can be rewritten as $G_0(\zeta^{-1}(y;-1)) \leq G_1(y)$ for all y. Lemma EX then yields an $F \in \mathcal{F}^\star$ by taking $\eta = -1$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$. Q.E.D.

Proof of Proposition 8. Suppose that the model is not misspecified. By Proposition 7, this implies that $\bar{\eta} \geq -1$ is well-defined. Fix any deterministic $\eta' \in [-1, \bar{\eta}]$, which satisfies $\mathbb{P}[\zeta(y_0;\eta') \leq y] \leq G_1(y)$ for all $y \geq \bar{y}$ due to the definition of $\bar{\eta}$. Applying Lemma EX with $\eta = \eta'$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$ shows that there exists an $F \in \mathcal{F}^\star$ with deterministic $\eta = \eta'$. Under this F, all conditional and unconditional averages of η equal η'. Because $\eta' \in [-1, \bar{\eta}]$ was arbitrary, we have shown that the sharp identified sets for $\mathbb{E}[\eta]$, $\mathbb{E}[\eta|B=0]$, and $\mathbb{E}[\eta|B=1]$ contain $[-1, \bar{\eta}]$, which was the claim. Q.E.D.

Proof of Proposition 9. We begin by showing that the bounds are valid. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Let $v_1 \equiv \nu(v_0, \epsilon, \eta) \equiv \zeta(v_0;\eta) - \epsilon\tau$.

Recall that $\zeta(v_0;\eta)$ is a decreasing function of η with $\zeta(v_0;0) = v_0$. Using this property gives

$$\mathbb{E}[\epsilon] = \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(v_0;\eta)\right] - \mathbb{E}\left[v_1\right]\right) \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;0)] - \mathbb{E}[v_1]\right) = \frac{1}{\tau}\left(\mathbb{E}[v_0] - \mathbb{E}[v_1]\right). \quad (32)$$

Lemma IDT implies that $\mathbb{E}[v_1] \leq \mathbb{E}[y_1]$, which gives (P9-LB). The conditional lower bounds (P9-LB-B1) and (P9-LB-B0) follow by using the same argument—observing that $v_1 \leq B\bar{y} + (1-B)v_1$, where $B = \mathbb{1}[v_1 \leq \bar{y}]$—then applying Lemma HM:

$$\mathbb{E}[\epsilon|B=0] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;0) - v_1|B=0]\right) \geq \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \leq q_0(1-\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right)$$
$$\text{and}\quad \mathbb{E}[\epsilon|B=1] \geq \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;0) - v_1|B=1]\right) \geq \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \leq q_0(\bar{p})] - \bar{y}\right).$$

The upper bound on the non-bunchers, (P9-UB-B0), follows by using the bound $\zeta(v_0;\eta) \leq \zeta(v_0;-1)$ and then applying Lemma HM in the other direction:

$$\begin{aligned}&\mathbb{E}[\epsilon|B=0]\\ &\quad\leq \frac{1}{\tau}\left(\mathbb{E}[\zeta(v_0;-1) - v_1|B=0]\right) \leq \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;-1)|y_0 > q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}]\right).\end{aligned}$$

Now we establish sharpness. We have assumed that the model is not misspecified, which Proposition 7 shows is equivalent to assuming that $\mathbb{P}[\zeta(y_0;-1) \leq y] = G_0(\zeta^{-1}(y;-1)) \leq G_1(y)$ for all $y \geq \bar{y}$. The assumptions of Lemma EX are therefore satisfied with $\eta = -1$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$.

For any $e \geq 0$, Lemma EX yields an $F \in \mathcal{F}^\star_{\text{D}} \subseteq \mathcal{F}^\star$ with deterministic income effects $\eta = -1$ such that (EX-ϵ-$\mathbb{E}$) holds. Taking $e = 0$ in (EX-ϵ-$\mathbb{E}$) produces the bound (P9-UB-B0). Taking e to be arbitrarily large leads to arbitrarily large bounds on $\mathbb{E}[\epsilon|B = 1]$ and $\mathbb{E}[\epsilon]$, implying that sharp bounds on these objects are infinite.

If we additionally assume $\bar{\eta} = 0$, then $\mathbb{P}[\zeta(y_0; 0) \leq y] = G_0(\zeta^{-1}(y; 0)) \leq G_1(y)$ for all $y \geq \bar{y}$. Lemma EX therefore also applies with $\eta = 0$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$. Taking $e = 0$ then yields (P9-LB-B1), and the law of iterated expectations gives (P9-LB). Q.E.D.

Proof of Proposition 10. We begin by showing that the bounds are valid. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Let $v_1 \equiv \nu(v_0, \epsilon, \eta)$.

Using (UE), we have

$$\mathbb{E}[\epsilon^u] = \frac{1}{\tau} \left(\mathbb{E}\left[\zeta^u(v_0; \eta) - v_1\right] \right) \geq \frac{1}{\tau} \left(\mathbb{E}\left[\zeta(v_0; -1)\right] - \tau - \mathbb{E}[v_1] \right),$$

where the inequality uses that $\zeta^u(v_0; \eta)$ is increasing in η, which implies that $\zeta^u(v_0; \eta) \geq \zeta^u(v_0; -1) = \zeta(v_0; -1) - \tau$. Applying Lemma IDT then gives

$$\mathbb{E}[\epsilon^u] \geq \frac{1}{\tau} \left(\mathbb{E}\left[\zeta(y_0; -1)\right] - \tau - \mathbb{E}[y_1] \right),$$

which is (P10-LB). Using the monotonicity of $\zeta^u(v_0; \eta)$ in η, then applying Lemmas HM and IDT, we obtain

$$\begin{aligned}
&\mathbb{E}[\epsilon^u|B = 1] \\
&\quad = \frac{1}{\tau} \left(\mathbb{E}[\zeta^u(v_0; \eta)|B = 1] - \mathbb{E}[v_1|B = 1] \right) \\
&\quad \geq \frac{1}{\tau} \left(\mathbb{E}[\zeta(v_0; -1)|B = 1] - \tau - \mathbb{E}[v_1|B = 1] \right) \\
&\quad \geq \frac{1}{\tau} \left(\mathbb{E}[\zeta(v_0; -1)|v_0 \leq q_0(\bar{p})] - \tau - \mathbb{E}[v_1|B = 1] \right) \\
&\quad \geq \frac{1}{\tau} \left(\mathbb{E}[\zeta(y_0; -1)|y_0 \leq q_0(\bar{p})] - \tau - \bar{y} \right),
\end{aligned}$$

which is (P10-LB-B1). Lemma HM also implies that

$$\begin{aligned}
&\mathbb{E}[\epsilon^u|B = 0] \\
&\quad \leq \frac{1}{\tau} \left(\mathbb{E}[\zeta^u(v_0; 0)|B = 0] - \mathbb{E}[v_1|B = 0] \right) \\
&\quad \leq \frac{1}{\tau} \left(\mathbb{E}[v_0|v_0 > q_0(\bar{p})] - \mathbb{E}[v_1|v_1 > \bar{y}] \right) = \frac{1}{\tau} \left(\mathbb{E}[y_0|y_0 > q_0(\bar{p})] - \mathbb{E}[y_1|y_1 > \bar{y}] \right),
\end{aligned}$$

which is (P10-UB-B0). To establish (P10-LB-B0), we first bound $\zeta^u(v_0; \eta)$ below by

$\zeta^u(v_0; -1) \equiv \zeta(v_0; -1) - \tau$:

$$\mathbb{E}[\epsilon^u | B = 0] \geq \frac{1}{\tau} \left(\mathbb{E}\left[\zeta(v_0; -1) | B = 0\right] - \tau - \mathbb{E}\left[v_1 | v_1 > \bar{y}\right] \right). \tag{33}$$

Finally, (28)–(29) in the proof of Proposition 5 show that $\Phi_{0|0}(v) \equiv \mathbb{P}[v_0 \leq v | B = 0] \leq \overline{G}(v)$ for all v, so

$$\mathbb{E}[\zeta(v_0; -1) | B = 0] \geq \int \zeta(y; -1)\, d\overline{G}(y). \tag{34}$$

Substituting (34) into (33) and using $\mathbb{E}[v_1 | v_1 > \bar{y}] = \mathbb{E}[y_1 | y_1 > \bar{y}]$ produces (P10-LB-B0).

We now establish sharpness by applying Lemma EX. Suppose that the model is not misspecified, which Proposition 7 shows is equivalent to assuming that $\mathbb{P}[\zeta(y_0; -1) \leq y] = G_0(\zeta^{-1}(y; -1)) \leq G_1(y)$ for all $y \geq \bar{y}$. The hypotheses of Lemma EX are therefore satisfied with $\eta = -1$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$.

For any $e \geq 0$, Lemma EX yields an $F \in \mathcal{F}^\star_{\text{D}} \subseteq \mathcal{F}^\star$ with deterministic income effects $\eta = -1$ such that (EX-ϵ-$\mathbb{E}$) holds. Taking $e = 0$ produces (P10-LB) and (P10-LB-B1). Taking e arbitrarily large shows that the corresponding sharp upper bounds are infinite. In Appendix SA.4, we show that (P10-LB-B0) can be attained under the single-crossing condition with $\eta = -1$.

If we additionally assume that $\bar{\eta} = 0$, then $\mathbb{P}[\zeta(y_0; 0) \leq y] = G_0(\zeta^{-1}(y; 0)) \leq G_1(y)$ for all $y \geq \bar{y}$. In this case, Lemma EX also applies with $\eta = 0$, $\Phi_0 = G_0$, and $\Phi_1 = G_1$. Taking $e = 0$ then produces $F \in \mathcal{F}^\star_{\text{D}} \subseteq \mathcal{F}^\star$ with deterministic $\eta = 0$ that attains (P10-UB-B0). Q.E.D.

Proof of Proposition 11. We begin by showing the bounds are valid. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Let $v_1 = \nu(v_0, \epsilon, \eta) \equiv \zeta(v_0; \eta) - \epsilon\tau$.

The lower bound (P11-LB-ϵ) on $\mathbb{E}[\epsilon]$ is the same as (P9-LB) in Proposition 9, which maintained $\mathcal{F}^\dagger = \mathcal{F}$, so is still valid when imposing Condition B, which makes $\mathcal{F}^\dagger$ a smaller set. To derive (P11-LB-ϵ^u) begin by writing

$$\begin{aligned}
\mathbb{E}[\epsilon^u] &= \frac{1}{\tau}\, \mathbb{E}[\zeta^u(v_0; \eta) - v_1] \\
&= \frac{1}{\tau}(\mathbb{E}[v_0 + \mathbb{E}[\eta | B, v_0](\tau - \pi(v_0))] - \mathbb{E}[v_1]) \\
&= \frac{1}{\tau}(\mathbb{E}[\zeta^u(v_0; \mathbb{E}[\eta | B, v_0])] - \mathbb{E}[v_1]) \geq \frac{1}{\tau}(\mathbb{E}[\zeta(v_0; \eta_{\text{MIN}}) + \tau\eta_{\text{MIN}}] - \mathbb{E}[v_1]),
\end{aligned} \tag{35}$$

where the first equality used the law of iterated expectations and the inequality used Condition B together with the monotonicity of $\zeta^u(v_0; \eta)$ in η. Lemma IDT with $F \in \mathcal{F}^\star$

then implies that

$$\frac{1}{\tau}(\mathbb{E}[\zeta(v_0;\eta_{\text{MIN}})+\tau\eta_{\text{MIN}}]-\mathbb{E}[v_1]) \geq \frac{1}{\tau}(\mathbb{E}[\zeta(y_0;\eta_{\text{MIN}})+\tau\eta_{\text{MIN}}]-\mathbb{E}[y_1]),$$

which is (P11-LB-ϵ^u).

To derive (P11-UB-ϵ) use the law of iterated expectations with Condition B:

$$\begin{aligned}
\mathbb{E}[\epsilon] &= \mathbb{E}[\epsilon|B=1]\bar{p}+\mathbb{E}[\epsilon|B=0](1-\bar{p}) \\
&\leq \epsilon_{\text{MAX}}\bar{p}+\frac{1}{\tau}\left(\mathbb{E}\left[\zeta(v_0;\eta)|B=0\right]-\mathbb{E}\left[v_1|B=0\right]\right)(1-\bar{p}) \\
&= \epsilon_{\text{MAX}}\bar{p}+\frac{1}{\tau}\left(\mathbb{E}\left[v_0-\mathbb{E}[\eta|B=0,v_0]\pi(v_0)|B=0\right]-\mathbb{E}\left[v_1|B=0\right]\right)(1-\bar{p}) \\
&\leq \epsilon_{\text{MAX}}\bar{p}+\frac{1}{\tau}\left(\mathbb{E}\left[\zeta(v_0;\eta_{\text{MIN}})|B=0\right]-\mathbb{E}\left[v_1|B=0\right]\right)(1-\bar{p}).
\end{aligned} \tag{36}$$

Lemmas HM and IDT with $F \in \mathcal{F}^\star$ imply

$$\mathbb{E}\left[\zeta(v_0;\eta_{\text{MIN}})|B=0\right]-\mathbb{E}\left[v_1|B=0\right] \leq \mathbb{E}\left[\zeta(y_0;\eta_{\text{MIN}})|y_0 \geq q_0(\bar{p})\right]-\mathbb{E}\left[y_1|y_1>\bar{y}\right],$$

which produces (P11-UB-ϵ) upon substitution into (36). A similar argument leads to (P11-UB-ϵ^u):

$$\begin{aligned}
\mathbb{E}[\epsilon^u] &= \mathbb{E}[\epsilon^u|B=1]\bar{p}+\mathbb{E}[\epsilon^u|B=0](1-\bar{p}) \\
&\leq \epsilon_{\text{MAX}}\bar{p}+\frac{1}{\tau}(\mathbb{E}[v_0|B=0]-\mathbb{E}[v_1|B=0])(1-\bar{p}),
\end{aligned} \tag{37}$$

where the inequality uses $\mathbb{E}[\epsilon^u|B=1] \leq \mathbb{E}[\epsilon|B=1] \leq \epsilon_{\text{MAX}}$ and $\zeta^u(v_0;\eta) \leq \zeta^u(v_0;0) = v_0$. Lemma HM and $F \in \mathcal{F}^\star$ imply

$$\mathbb{E}[v_0|B=0]-\mathbb{E}[v_1|B=0] \leq \mathbb{E}[y_0|y_0 \geq q_0(\bar{p})]-\mathbb{E}[y_1|y_1>\bar{y}].$$

Substituting into (37) produces (P11-UB-ϵ^u).

To establish sharpness, assume that the model is not misspecified. Then $\bar{\eta} \in [-1,0]$ exists and for any deterministic $\eta \leq \bar{\eta}$ we have $G_0(\zeta^{-1}(y;\eta)) \leq G_1(y)$ for all $y \geq \bar{y}$. Lemma EX can be applied with this choice of η, setting $\Phi_0 = G_0$ and $\Phi_1 = G_1$.

The conclusion of Lemma EX is that there exists an $F \in \mathcal{F}^\star_{\text{D}}$ for any $e \geq 0$. Because $\eta \geq \eta_{\text{MIN}}$, we know that such an F must satisfy $\mathbb{E}_F[\eta|v_0, B=b] \geq \eta_{\text{MIN}}$ for $b=0,1$ and almost every v_0. To show that $F \in \mathcal{F}^\star$, we need to verify that $\mathbb{E}_F[\epsilon|B=1] \leq \epsilon_{\text{MAX}}$. We do this for each bound separately, choosing an $\eta \in [\eta_{\text{MIN}}, \bar{\eta}]$ and $e \geq 0$ so that $F \in \mathcal{F}^\star$ also attains the claimed bound under the stated side conditions on how η_{MIN} and ϵ_{MAX} relate to $\bar{\eta}$ and $\underline{\epsilon}(\eta_{\text{MIN}})$.

- (P11-LB-ϵ): Suppose that $\bar{\eta} = 0$. Take $\eta = 0$ and $e = 0$. Then $\eta_{\text{MIN}} \leq \eta \leq \bar{\eta}$ and $e \geq 0$. Take an $F \in \mathcal{F}^{\star}_{\text{D}}$ produced by Lemma EX with these choices. Applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$) shows that this F produces (P11-LB-ϵ). It also satisfies $\mathbb{E}_F[\epsilon|B = 1] = \underline{\epsilon}(0)$, which is the lower bound (P9-LB-B1) from Proposition 9. Imposing Condition B cannot make this lower bound any smaller. Because $\mathcal{F}^{\star}$ is assumed to be non-empty, it follows that

$$\underline{\epsilon}(0) = \mathbb{E}_F[\epsilon|B = 1] \leq \epsilon_{\text{MAX}}, \tag{38}$$

which verifies that $F \in \mathcal{F}^{\star}$.

- (P11-LB-ϵ^u): Suppose that $\bar{\eta} \geq \eta_{\text{MIN}}$ and $\epsilon_{\text{MAX}} \geq \underline{\epsilon}(\eta_{\text{MIN}})$. Take $\eta = \eta_{\text{MIN}}$ and $e = 0$. Then $\eta_{\text{MIN}} \leq \eta \leq \bar{\eta}$ and $e \geq 0$. Take an $F \in \mathcal{F}^{\star}_{\text{D}}$ produced by Lemma EX with these choices. Applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$) then adding $\eta = \eta_{\text{MIN}}$ to create an uncompensated elasticity shows that this F produces (P11-LB-ϵ^u). It also satisfies $\mathbb{E}_F[\epsilon|B = 1] = \underline{\epsilon}(\eta_{\text{MIN}}) \leq \epsilon_{\text{MAX}}$, which verifies that $F \in \mathcal{F}^{\star}$.

- (P11-UB-ϵ): Suppose that $\bar{\eta} \geq \eta_{\text{MIN}}$ and $\epsilon_{\text{MAX}} \geq \underline{\epsilon}(\eta_{\text{MIN}})$. Take $\eta = \eta_{\text{MIN}}$, which satisfies $\eta_{\text{MIN}} \leq \eta \leq \bar{\eta}$. Take

$$\begin{aligned} e &= \tau\epsilon_{\text{MAX}} + \bar{y} - \mathbb{E}[\zeta(y_0; \eta_{\text{MIN}})|y_0 \leq q_0(\bar{p})] \\ &\geq \tau\underline{\epsilon}(\eta_{\text{MIN}}) + \bar{y} - \mathbb{E}[\zeta(y_0; \eta_{\text{MIN}})|y_0 \leq q_0(\bar{p})] = 0, \end{aligned}$$

where the inequality follows from $\epsilon_{\text{MAX}} \geq \underline{\epsilon}(\eta_{\text{MIN}})$ and the definition of $\underline{\epsilon}(\eta_{\text{MIN}})$. Take an $F \in \mathcal{F}^{\star}_{\text{D}}$ produced by Lemma EX with these choices. Applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$) shows that this F produces (P11-UB-ϵ). It also satisfies $\mathbb{E}_F[\epsilon|B = 1] = \epsilon_{\text{MAX}}$, which verifies that $F \in \mathcal{F}^{\star}$.

- (P11-UB-ϵ^u): Suppose $\bar{\eta} = 0$. Take $\eta = 0$, which satisfies $\eta_{\text{MIN}} \leq \eta \leq \bar{\eta}$. Take

$$\begin{aligned} e &\equiv \tau\epsilon_{\text{MAX}} + \bar{y} - \mathbb{E}[\zeta(y_0; 0)|y_0 \leq q_0(\bar{p})] \\ &\geq \tau\underline{\epsilon}(0) + \bar{y} - \mathbb{E}[\zeta(y_0; 0)|y_0 \leq q_0(\bar{p})] = 0, \end{aligned}$$

where the inequality uses (38) and the definition of $\underline{\epsilon}(0)$. Take an $F \in \mathcal{F}^{\star}_{\text{D}}$ produced by Lemma EX with these choices. Applying the law of iterated expectations to (EX-ϵ-$\mathbb{E}$) with this F produces (P11-UB-ϵ^u), because compensated and uncompensated elasticities are the same when $\eta = 0$, and $\zeta(y_0; 0) = y_0$. This choice of F also satisfies $\mathbb{E}_F[\epsilon|B = 1] = \epsilon_{\text{MAX}}$, which verifies that $F \in \mathcal{F}^{\star}$.

Q.E.D.

SA.3 Proofs for Section 7

Our proofs in this section make use of the following lemma in combination with Lemmas EX and HM.

Lemma VX. (Extension of virtual earnings distributions) Suppose that $\eta \in [-1, 0]$ is deterministic. Suppose that $\mathbb{P}[\zeta(y_0; \eta) \leq y] \leq G_1(y)$ for all y such that $\bar{y} < y < \min\{\zeta(\bar{\bar{y}}_0; \eta), \bar{\bar{y}}_1\}$. Then there exist distribution functions Φ_0 and Φ_1 that satisfy the following properties:

VX1. $\Phi_0(y) = G_0(y)$ for all $y \in (\bar{y}, \bar{\bar{y}}_0)$.

VX2. $\Phi_1(y) = G_1(y)$ for all $y \in [\bar{y}, \bar{\bar{y}}_1)$.

VX3. Φ_0 and Φ_1 are strictly increasing and absolutely continuous for all $y > \bar{y}$.

VX4. $\Phi_0(\zeta^{-1}(y; \eta)) \leq \Phi_1(y)$ for all $y > \bar{y}$.

Proof of Lemma VX. For $d = 0, 1$ define

$$\Phi_d(y) \equiv \begin{cases} G_d(y), & \text{if } y < \bar{\bar{y}}_d \\ \Gamma_d(y), & \text{if } y \geq \bar{\bar{y}}_d \end{cases},$$

where $\Gamma_d(y)$ is a continuous extension of G_d for $y > \bar{\bar{y}}_d$ that makes Φ_d a strictly increasing, absolutely continuous distribution function. Then VX1, VX2, and VX3 are satisfied. To ensure VX4, we impose an additional condition on these extensions depending on where $\min\{\zeta(\bar{\bar{y}}_0; \eta), \bar{\bar{y}}_1\}$ is attained.

Suppose first that $\zeta(\bar{\bar{y}}_0; \eta) \leq \bar{\bar{y}}_1$. Let Γ_1 be any extension of G_1. Take Γ_0 to be such that $\Gamma_0(y) \leq \Phi_1(\zeta(y; \eta))$ for all $y \geq \bar{\bar{y}}_0$. This is possible because

$$\Gamma_0(\bar{\bar{y}}_0) = G_0(\bar{\bar{y}}_0) = \mathbb{P}[y_0 \leq \bar{\bar{y}}_0] = \mathbb{P}[\zeta(y_0; \eta) \leq \zeta(\bar{\bar{y}}_0; \eta)] \leq G_1(\zeta(\bar{\bar{y}}_0; \eta)) = \Phi_1(\zeta(\bar{\bar{y}}_0; \eta)),$$

where the first equality follows from continuity at $\bar{\bar{y}}_0$, the third equality follows because $\zeta(\cdot; \eta)$ is a strictly increasing function, and the inequality is by assumption, because $\zeta(\bar{\bar{y}}_0; \eta) \leq \bar{\bar{y}}_1$. We conclude that $\Phi_0(y) \leq \Phi_1(\zeta(y; \eta))$ for all $y \geq \bar{\bar{y}}_0$ or, equivalently, that $\Phi_0(\zeta^{-1}(y; \eta)) \leq \Phi_1(y)$ for all $y \geq \zeta(\bar{\bar{y}}_0; \eta)$. The same condition is satisfied by assumption for $y < \bar{\bar{y}}_0$, because $\Phi_0(y) = G_0(y)$ and $\Phi_1(\zeta(y; \eta)) = G_1(\zeta(y; \eta))$. We conclude that Φ_0 and Φ_1 also satisfy VX4.

On the other hand, if $\bar{\bar{y}}_1 \leq \zeta(\bar{\bar{y}}_0; \eta)$, then we take Γ_0 to be any extension of G_0 and choose Γ_1 to be such that $\Gamma_1(y) \geq \Phi_0(\zeta^{-1}(y; \eta))$ for all $y \geq \bar{\bar{y}}_1$. This is possible because

$$\Gamma_1(\bar{\bar{y}}_1) = G_1(\bar{\bar{y}}_1) \geq \mathbb{P}[\zeta(y_0; \eta) \leq \bar{\bar{y}}_1] = G_0(\zeta^{-1}(\bar{\bar{y}}_1; \eta)) = \Phi_0(\zeta^{-1}(\bar{\bar{y}}_1; \eta)),$$

where the inequality is by assumption for $y \leq \bar{\bar{y}}_1$ and the final equality follows from $\zeta^{-1}(\bar{\bar{y}}_1;\eta) \leq \zeta^{-1}(\zeta(\bar{\bar{y}}_0;\eta);\eta) = \bar{\bar{y}}_0$. We conclude that $\Phi_1(y) \geq \Phi_0(\zeta^{-1}(y;\eta))$ for $y \geq \bar{\bar{y}}_1$. The same condition is satisfied by assumption for $y < \bar{\bar{y}}_1$, because $\Phi_1(y) = G_1(y) \geq G_0(\zeta^{-1}(y;\eta)) = \Phi_0(\zeta^{-1}(y;\eta))$, noting that $\zeta^{-1}(y;\eta) \leq \zeta^{-1}(\bar{\bar{y}}_1;\eta) \leq \zeta^{-1}(\zeta(\bar{\bar{y}}_0;\eta);\eta) = \bar{\bar{y}}_0$. We conclude that Φ_0 and Φ_1 also satisfy VX4. Q.E.D.

Proof of Proposition 12. Suppose that the model is not misspecified, so that $\mathcal{F}^\star$ is non-empty. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Let $v_1 \equiv \nu(v_0, \epsilon, \eta) \equiv \zeta(v_0;\eta) - \epsilon\tau$. Suppose that $y \in (\bar{y}, \min\{\bar{\bar{y}}_1, \zeta(\bar{\bar{y}}_0;-1)\})$. Then

$$G_1(y) = \mathbb{P}_F[v_1 \leq y] \geq \mathbb{P}_F[\zeta(v_0;-1) \leq y] = \mathbb{P}_F[v_0 \leq \zeta^{-1}(y;-1)] = G_0(\zeta^{-1}(y;-1)),$$

where the first and last equalities used $F \in \mathcal{F}^\star$—noting that $y \leq \zeta(\bar{\bar{y}}_0;-1)$ if and only if $\zeta^{-1}(y;-1) \leq \bar{\bar{y}}_0$—and the inequality follows because $\epsilon \geq 0$ and $\zeta(v_0;\eta)$ is a decreasing function of η. The claimed inequality follows after rewriting $G_0(\zeta^{-1}(y;-1)) = \mathbb{P}[\zeta(y_0;-1) \leq y]$.

To show the converse, suppose that $\mathbb{P}[\zeta(y_0;-1) \leq y] = G_0(\zeta^{-1}(y;-1)) \leq G_1(y)$ for all y such that $\bar{y} < y < \min\{\bar{\bar{y}}_1, \zeta(\bar{\bar{y}}_0;-1)\}$. Lemma VX shows that if this condition holds, then there exist distribution functions Φ_0 and Φ_1 such that $\Phi_0(\zeta^{-1}(y;-1)) \leq \Phi_1(y)$ for all $y > \bar{y}$. Applying Lemma EX with these Φ_0 and Φ_1 shows that there exists an $F \in \mathcal{F}^\star$, so that the model is not misspecified. Q.E.D.

Proof of Proposition 13. Let $F \in \mathcal{F}^\star$ and let (ϵ, η, v_0) be random variables with distribution F. Let $v_1 \equiv \nu(v_0, \epsilon, \eta) \equiv \zeta(v_0;\eta) - \epsilon\tau$.

We first show that (P13-LB-ϵ) is valid. Because $\zeta(v_0;\eta)$ is decreasing in η with $\zeta(v_0;0) = v_0$,

$$\begin{aligned}\mathbb{E}^\star[\epsilon] &\geq \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0;\eta)] - \mathbb{E}^\star[v_1]\right)\\ &\geq \frac{1}{\tau}\left(\mathbb{E}^\star[v_0] - \mathbb{E}^\star[v_1]\right) = \frac{1}{\tau}\left(\mathbb{E}[y_0|y_0 \in \mathcal{Y}_0^\star] - \mathbb{E}[v_1|v_0 \in \mathcal{Y}_0^\star]\right),\end{aligned} \tag{39}$$

where the final equality follows because $F \in \mathcal{F}^\star$. We derive a lower bound on the final term by noting that $v_0 \in \mathcal{Y}_0^\star$ implies $v_1 \leq \zeta(y_0^\star;-1) \equiv y_1^\star$, then trimming the mass for the event $v_0 \in \mathcal{Y}_0^\star$:

$$\mathbb{E}[v_1|v_0 \in \mathcal{Y}_0^\star] = \mathbb{E}[v_1|v_0 \in \mathcal{Y}_0^\star, v_1 \leq y_1^\star] \leq \mathbb{E}[y_1|q_1^\star(1 - p_{0|1}^\star) < y_1 \leq y_1^\star], \tag{40}$$

where $p_{0|1}^\star \equiv \mathbb{P}[v_0 \in \mathcal{Y}_0^\star|v_1 \leq y_1^\star]$, $q_1^\star$ is the quantile function of y_1, conditional on

$y_1 \leq y_1^\star$, and we used $F \in \mathcal{F}^\star$ in the inequality.[17] Notice that because $q_1^\star(\alpha) \leq y_1^\star$ for any α and because $F \in \mathcal{F}^\star$,

$$\begin{aligned}
\mathbb{P}\left[y_1 \leq q_1^\star(1 - p_{0|1}^\star)\right] &= \mathbb{P}\left[y_1 \leq q_1^\star(1 - p_{0|1}^\star) | y_1 \leq y_1^\star\right] \mathbb{P}[y_1 \leq y_1^\star] \\
&= (1 - p_{0|1}^\star)\, \mathbb{P}[v_1 \leq y_1^\star] \\
&= \mathbb{P}[v_1 \leq y_1^\star] - \mathbb{P}[v_0 \in \mathcal{Y}_0^\star, v_1 \leq y_1^\star] \\
&= \mathbb{P}[v_1 \leq y_1^\star] - \mathbb{P}[v_0 \in \mathcal{Y}_0^\star] \\
&= G_1(y_1^\star) - G_0(y_0^\star) \equiv p_1^\star - p_0^\star.
\end{aligned}$$

We conclude that $q_1^\star(1 - p_{0|1}^\star) = q_1(p_1^\star - p_0^\star) \equiv q_1^{\text{L}}$, so that

$$\mathbb{E}[v_1 | y_0 \in \mathcal{Y}_0^\star] \leq \mathbb{E}[y_1 | q_1^{\text{L}} < y_1 \leq y_1^\star]. \tag{41}$$

Using (41) in (39) produces (P13-LB-ϵ). The lower bound for the average uncompensated elasticity follows from a similar argument:

$$\begin{aligned}
\mathbb{E}^\star[\epsilon^u] &= \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0; \eta)] + \tau\eta - \mathbb{E}^\star[v_1]\right) \\
&\geq \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0; \eta_{\text{MIN}})] + \tau\eta_{\text{MIN}} - \mathbb{E}^\star[v_1]\right) \\
&= \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0; \eta_{\text{MIN}}) | y_0 \in \mathcal{Y}_0^\star] + \tau\eta_{\text{MIN}} - \mathbb{E}[v_1 | v_0 \in \mathcal{Y}_0^\star]\right).
\end{aligned}$$

Using (41) produces (P13-LB-ϵ^u).

Next, we show that the upper bounds (P13-UB-ϵ) and (P13-UB-ϵ^u) are valid, which is a more involved argument. The law of iterated expectations combined with Condition B* gives

$$\mathbb{E}^\star[\epsilon] \leq \epsilon_{\text{MAX}}\bar{p}^\star + \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0; \eta_{\text{MIN}}) | v_1 > \bar{y}] - \mathbb{E}^\star[v_1 | v_1 > \bar{y}]\right)(1 - \bar{p}^\star), \tag{42}$$

$$\text{and} \quad \mathbb{E}^\star[\epsilon^u] \leq \epsilon_{\text{MAX}}\bar{p}^\star + \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0; 0) | v_1 > \bar{y}] - \mathbb{E}^\star[v_1 | v_1 > \bar{y}]\right)(1 - \bar{p}^\star), \tag{43}$$

where $\bar{p}^\star \equiv \mathbb{P}[v_1 \leq \bar{y} | v_0 \in \mathcal{Y}_0^\star]$. We bound each of the two conditional expectations in (42)–(43) in turn. The arguments for (42) and (43) are identical, except for the deterministic value of η at which the ZER is evaluated. The following argument uses a generic $\tilde{\eta}$ that we evaluate for $\tilde{\eta} \in \{\eta_{\text{MIN}}, 0\}$ at the end.

For the first conditional expectation in (42)–(43), we derive an upper bound by

[17] Note that the strict inequality $q_1^\star(1 - p_0^\star) < y_1$ is needed because y_1 can have a point mass at $\bar{y}$.

trimming the mass in the event $v_1 > \bar{y}$:

$$\mathbb{E}^\star[\zeta(v_0;\tilde{\eta})|v_1 > \bar{y}] \equiv \mathbb{E}[\zeta(v_0;\tilde{\eta})|v_0 \in \mathcal{Y}_0^\star, v_1 > \bar{y}] \leq \mathbb{E}[\zeta(y_0;\tilde{\eta})|q_0^\star(\bar{p}^\star) \leq y_0 \leq y_0^\star], \quad (44)$$

where $q_0^\star$ is the quantile function of y_0, conditional on $y_0 \leq y_0^\star$, and we used $F \in \mathcal{F}^\star$ in the inequality. Notice that for any quantile α,

$$\mathbb{P}[y_0 \leq q_0^\star(\alpha)] = \mathbb{P}[y_0 \leq q_0^\star(\alpha)|v_0 \in \mathcal{Y}_0^\star]\,\mathbb{P}[v_0 \in \mathcal{Y}_0^\star] = \alpha p_0^\star,$$

with $p_0^\star \equiv \mathbb{P}[y_0 \in \mathcal{Y}_0^\star] = \mathbb{P}[v_0 \in \mathcal{Y}_0^\star]$ because $F \in \mathcal{F}^\star$. This implies that $q_0^\star(\alpha) = q_0(\alpha p_0^\star)$. In particular, $q_0^\star(\bar{p}^\star) = q_0(\bar{p}^\star p_0^\star) \leq q_0(\bar{p})$ because

$$\bar{p}^\star p_0^\star = \mathbb{P}[v_0 \in \mathcal{Y}_0^\star, v_1 \leq \bar{y}] \leq \mathbb{P}[v_1 \leq \bar{y}] \equiv \bar{p}. \quad (45)$$

Using this observation with the law of iterated expectations in (44) gives

$$\begin{aligned}
&\mathbb{E}^\star[\zeta(v_0;\tilde{\eta})|v_1 > \bar{y}] \\
&\quad \leq \mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}^\star p_0^\star) \leq y_0 \leq q_0(\bar{p})\right]\left(\frac{\bar{p}-\bar{p}^\star p_0^\star}{p_0^\star - \bar{p}^\star p_0^\star}\right) \\
&\qquad + \mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \leq y_0 \leq y_0^\star\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star - \bar{p}^\star p_0^\star}\right) \\
&\quad \leq \zeta(q_0(\bar{p});-1)\left(\frac{\bar{p}-\bar{p}^\star p_0^\star}{p_0^\star - \bar{p}^\star p_0^\star}\right) + \mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \leq y_0 \leq y_0^\star\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star - \bar{p}^\star p_0^\star}\right) \\
&\quad \leq \left[(\tau\epsilon_{\text{MAX}} + \bar{y})\left(\frac{\bar{p}-\bar{p}^\star p_0^\star}{p_0^\star}\right) + \mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \leq y_0 \leq y_0^\star\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right)\right]\frac{1}{1-\bar{p}^\star}, \quad (46)
\end{aligned}$$

where the final inequality used Condition B* and factored out $1/(1-\bar{p}^\star)$.

For the second term of (42), we derive a lower bound by applying a similar trimming argument:

$$\mathbb{E}^\star[v_1|v_1 > \bar{y}] \equiv \mathbb{E}[v_1|v_0 \in \mathcal{Y}_0^\star, v_1 > \bar{y}] \geq \mathbb{E}\left[y_1 \middle| \bar{y} < y_1 \leq \bar{q}_1\left(\frac{p_0^\star - \bar{p}^\star p_0^\star}{1-\bar{p}}\right)\right], \quad (47)$$

where the trimming used the relationship

$$\begin{aligned}
\mathbb{P}[v_0 \in \mathcal{Y}_0^\star|v_1 > \bar{y}] &= \frac{\mathbb{P}[v_0 \in \mathcal{Y}_0^\star, v_1 > \bar{y}]}{\mathbb{P}[v_1 > \bar{y}]} \\
&= \frac{\mathbb{P}[v_0 \in \mathcal{Y}_0^\star] - \mathbb{P}[v_0 \in \mathcal{Y}_0^\star, v_1 \leq \bar{y}]}{\mathbb{P}[v_1 > \bar{y}]} \equiv \frac{p_0^\star - \bar{p}^\star p_0^\star}{1-\bar{p}}.
\end{aligned}$$

Notice that for any quantile α,

$$\mathbb{P}[y_1 \le \bar{q}_1(\alpha)] = \mathbb{P}[y_1 \le \bar{q}_1(\alpha)|y_1 \le \bar{y}]\bar{p} + \mathbb{P}[y_1 \le \bar{q}_1(\alpha)|y_1 > \bar{y}](1-\bar{p}) = \bar{p} + \alpha(1-\bar{p}),$$

which combined with (47) implies that

$$\mathbb{E}^\star[v_1|v_1 > \bar{y}] \le \mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1\left(p_0^\star + \bar{p} - \bar{p}^\star p_0^\star\right)\right].$$

We split this conditional expectation into two terms—recalling from (45) that $\bar{p}^\star p_0^\star \le \bar{p}$—then bound one part below by $\bar{y}$:

$$\begin{aligned}\mathbb{E}^\star[v_1|v_1 > \bar{y}] &\ge \mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1(p_0^\star)\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star - \bar{p}^\star p_0^\star}\right) + \bar{y}\left(\frac{\bar{p} - \bar{p}^\star p_0^\star}{p_0^\star - \bar{p}^\star p_0^\star}\right)\\ &= \left[\mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1(p_0^\star)\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right) + \bar{y}\left(\frac{\bar{p} - \bar{p}^\star p_0^\star}{p_0^\star}\right)\right]\frac{1}{1-\bar{p}^\star}. \qquad (48)\end{aligned}$$

We now return to (42), substituting (46) and (48):

$$\begin{aligned}\mathbb{E}^\star[\epsilon] &\le \epsilon_{\text{MAX}}\bar{p}^\star + \frac{1}{\tau}\left(\tau\epsilon_{\text{MAX}} + \bar{y}\right)\left(\frac{\bar{p} - \bar{p}^\star p_0^\star}{p_0^\star}\right)\\ &\quad + \frac{1}{\tau}\,\mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \le y_0 \le y_0^\star\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right)\\ &\quad - \frac{1}{\tau}\,\mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1(p_0^\star)\right]\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right)\\ &\quad - \frac{1}{\tau}\bar{y}\left(\frac{\bar{p} - \bar{p}^\star p_0^\star}{p_0^\star}\right)\\ &= \epsilon_{\text{MAX}}\left(\bar{p}^\star + \frac{\bar{p} - \bar{p}^\star p_0^\star}{p_0^\star}\right)\\ &\quad + \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \le y_0 \le y_0^\star\right] - \mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1(p_0^\star)\right]\right)\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right)\\ &= \frac{\bar{p}\epsilon_{\text{MAX}}}{p_0^\star} + \frac{1}{\tau}\left(\mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) \le y_0 \le y_0^\star\right] - \mathbb{E}\left[y_1|\bar{y} < y_1 \le q_1(p_0^\star)\right]\right)\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right),\end{aligned}$$

which is (P13-UB-ϵ) when evaluated at $\tilde{\eta} = \eta_{\text{MIN}}$. Similarly, substituting (46) and (48) evaluated at $\tilde{\eta} = 0$ into (43) produces (P13-UB-ϵ^u) because $\zeta(y_0; 0) = y_0$.

To establish sharpness, assume that the model is not misspecified. Then $\bar{\eta}^\star \in [-1, 0]$ exists and for any deterministic $\tilde{\eta} \le \bar{\eta}^\star$, Proposition 12 implies that $G_0(\zeta^{-1}(y;\tilde{\eta})) \le G_1(y)$ for all $y \in (\bar{y}, \min\{\zeta(\bar{\bar{y}}_0;\tilde{\eta}), \bar{\bar{y}}_1\})$. Appealing to Lemma VX yields distribution functions Φ_0 and Φ_1 such that $\Phi_0(\zeta^{-1}(y;\tilde{\eta})) \le \Phi_1(y)$ for all $y > \bar{y}$.

Next, we apply Lemma EX with Φ_0, Φ_1, $\eta = \tilde{\eta}$ deterministic, and

$$\begin{aligned} e &\equiv \tau\epsilon_{\text{MAX}} + (\bar{y} - \mathbb{E}[\zeta(y_0;\tilde{\eta})|y_0 \leq q_0(\bar{p})]) \\ &\geq (\zeta(q_0(\bar{p}); -1) - \bar{y}) + (\bar{y} - \mathbb{E}[\zeta(y_0;\tilde{\eta})|y_0 \leq q_0(\bar{p})]) \\ &\geq \mathbb{E}[\zeta(y_0;-1)|y_0 \leq q_0(\bar{p})] - \mathbb{E}[\zeta(y_0;\tilde{\eta})|y_0 \leq q_0(\bar{p})] \geq 0, \end{aligned} \tag{49}$$

where the first inequality used Condition B*. The conclusion of Lemma EX is that there exists an $F \in \mathcal{F}_{\text{D}}^\star$ with ϵ defined by (EX-ϵ), so that

$$\begin{aligned} \mathbb{E}_F^\star[\epsilon|B=1] &\equiv \mathbb{E}_F[\epsilon|B=1, v_0 \in \mathcal{Y}_0^\star] \\ &= \mathbb{E}_F[\epsilon|B=1] \\ &= \frac{1}{\tau}(\mathbb{E}[\zeta(y_0;\tilde{\eta})|y_0 \leq q_0(\bar{p})] - \bar{y} + e) = \epsilon_{\text{MAX}}, \end{aligned} \tag{50}$$

where the second equality follows from (EX-ϵ), the third equality is (EX-ϵ) with $F \in \mathcal{F}_{\text{D}}^\star$, and the final equality substitutes the choice of e in (49). From (EX-ϵ-$\mathbb{E}$) we also obtain

$$\mathbb{E}_F^\star[\epsilon|B=0] = \frac{1}{\tau}\big(\,\mathbb{E}\left[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) < y_0 \leq y_0^\star\right] - \mathbb{E}[y_1|\bar{y} < y_1 \leq q_1^{\text{U}}]\big),$$

as well as

$$\mathbb{P}_F^\star[B=1] \equiv \mathbb{P}_F[B=1|v_0 \in \mathcal{Y}_0^\star] = \frac{\mathbb{P}[B=1, v_0 \in \mathcal{Y}_0^\star]}{\mathbb{P}[v_0 \in \mathcal{Y}_0^\star]} = \frac{\bar{p}}{p_0^\star}.$$

Applying the law of iterated expectations, we conclude that

$$\mathbb{E}_F^\star[\epsilon] = \frac{\bar{p}\epsilon_{\text{MAX}}}{p_0^\star} + \frac{1}{\tau}\big(\,\mathbb{E}[\zeta(y_0;\tilde{\eta})|q_0(\bar{p}) < y_0 \leq y_0^\star] - \mathbb{E}[y_1|\bar{y} < y_1 \leq q_1^{\text{U}}]\big)\left(\frac{p_0^\star - \bar{p}}{p_0^\star}\right). \tag{51}$$

Now we evaluate (51) for different choices of $\tilde{\eta}$. If $\bar{\eta}^\star \geq \eta_{\text{MIN}}$, then we can take $\tilde{\eta} = \eta_{\text{MIN}}$ to obtain (P13-UB-ϵ). If $\bar{\eta}^\star = 0$, then we can take $\eta = 0$ to obtain (P13-UB-ϵ^u), noting that $\zeta(y_0;0) = y_0$ and $\mathbb{E}_F^\star[\epsilon^u] = \mathbb{E}_F^\star[\epsilon]$ in the absence of income effects. To show that $F \in \mathcal{F}^\star$ in either case, we still need to verify that $\mathbb{E}_F^\star[\epsilon|B=1] \leq \epsilon_{\text{MAX}}$ and that $\mathbb{E}_F[\eta|v_0, B=b] \geq \eta_{\text{MIN}}$ for $b = 0, 1$ and almost every $v_0 \leq y_0^\star$. The first condition was shown to be satisfied in (50). The second condition is satisfied because in either case we set $\eta = \tilde{\eta}$ to be deterministic with $\tilde{\eta} \geq \eta_{\text{MIN}}$.

The same arguments apply with minor changes if we additionally maintain Condition D, so that $\eta = \tilde{\eta}$ is deterministic. In particular, redefine $y_1^\star \equiv \zeta(y_0^\star;\tilde{\eta})$, replace

(39) with

$$\mathbb{E}^\star[\epsilon] = \frac{1}{\tau}\left(\mathbb{E}[\zeta(y_0;\tilde{\eta})|y_0 \in \mathcal{Y}_0^\star] - \mathbb{E}[v_1|v_0 \in \mathcal{Y}_0^\star]\right),$$

and replacing (42) with

$$\mathbb{E}^\star[\epsilon] \leq \epsilon_{\text{MAX}}\bar{p}^\star + \frac{1}{\tau}\left(\mathbb{E}^\star[\zeta(v_0;\tilde{\eta})|v_1 > \bar{y}] - \mathbb{E}^\star[v_1|v_1 > \bar{y}]\right)(1-\bar{p}^\star). \tag{52}$$

The rest of the proof of both validity and sharpness then follow the same steps as when η is heterogeneous. Q.E.D.

SA.4 Sharp lower bounds on non-buncher elasticities

Proposition 5 provided the sharp lower bound on the average compensated elasticity for non-bunchers, $\mathbb{E}[\epsilon|B=0]$, under a side condition that $G'_{1|0}$ and $G'_0/(1-\bar{p})$ only cross once. In this appendix, we extend the result to characterize the sharp bound when these two functions cross a finite number of times. As in Proposition 5, $\mathcal{F}^\star$ is defined by Conditions T and D. We assume throughout that $\mathcal{F}^\star$ is non-empty so that the model is not misspecified.

We start by transforming the problem into something more tractable. Each $F \in \mathcal{F}^\star$ implies an average compensated elasticity for the non-bunchers:

$$\gamma(F) \equiv \mathbb{E}_F\left[\epsilon|\nu(y_0,\epsilon,\eta) > \bar{y}\right] = \frac{1}{\tau}\left(\mathbb{E}_F\left[z_0|y_1 > \bar{y}\right] - \mathbb{E}\left[y_1|y_1 > \bar{y}\right]\right),$$

The sharp lower bound on this object is

$$\gamma_{\text{LB}} \equiv \inf\left\{\gamma(F) : F \in \mathcal{F}^\star\right\} = \frac{1}{\tau}\Big(\inf\left\{\mathbb{E}_F\left[z_0|y_1 > \bar{y}\right] : F \in \mathcal{F}^\star\right\} - \mathbb{E}\left[y_1|y_1 > \bar{y}\right]\Big).$$

This expression shows that characterizing the sharp lower bound requires finding a distribution of compensated elasticity F in the identified set $\mathcal{F}^\star$ that produces the smallest conditional average of the ZER, $z_0 \equiv \zeta(y_0;\eta)$, for the non-bunchers. In the next proposition, we show that this problem is equivalent to finding a distribution of the ZER that satisfies two restrictions.

Proposition SA.1. Suppose that $\eta \in [-1,0]$ is constant and let $z_0 \equiv \zeta(y_0;\eta)$. Let $\mathcal{H}$ denote the set of all absolutely continuous and strictly increasing distribution functions H such that $H(\bar{z}) = 0$, where $\bar{z} \equiv \zeta(\bar{y};\eta)$. Define the set

$$\mathcal{H}^\star \equiv \left\{H \in \mathcal{H} : H'(z) \leq \frac{h_0(z)}{1-\bar{p}} \equiv \bar{h}(z) \quad \text{and} \quad H(z) \leq G_{1|0}(z)\right\},$$

where $H'(z)$ is the first derivative of H, $h_0(z)$ is the density of z_0, and $G_{1|0}(z) \equiv \mathbb{P}[y_1 \leq z|y_1 > \bar{y}]$ is the distribution of system 1 earnings for the non-bunchers. Then $\mathcal{H}^\star$ is non-empty if and only if $\mathcal{F}^\star$ is non-empty. When this is the case,

$$\gamma_{\text{LB}} = \frac{1}{\tau}\left(\inf\left\{\int zH'(z)\,dz : H \in \mathcal{H}^\star\right\} - \mathbb{E}[y_1|y_1 > \bar{y}]\right).$$

Proof of Proposition SA.1. We will show that

$$\{\mathbb{E}_F[z_0|y_1 > \bar{y}] : F \in \mathcal{F}^\star\} = \left\{\int zH'(z)\,dz : H \in \mathcal{H}^\star\right\}, \tag{53}$$

which implies both that $\mathcal{H}^\star$ is non-empty exactly when $\mathcal{F}^\star$ is non-empty, and that the infima of these two sets are the same.

First, take any $F \in \mathcal{F}^\star$. Let ϵ be a random variable that has distribution F, conditional on y_0. Then define the virtual earnings

$$v \equiv y_0 - \eta\pi(y_0) - \tau\epsilon \equiv z_0 - \tau\epsilon.$$

Because $F \in \mathcal{F}^\star$, this virtual earnings function satisfies

$$\mathbb{P}_F[v > \bar{y}] = \mathbb{P}_F[y_1 > \bar{y}] = 1 - G_1(\bar{y}) \equiv 1 - \bar{p}.$$

Define the distribution function

$$H(z) \equiv \mathbb{P}_F[z_0 \leq z|v > \bar{y}],$$

which satisfies $H(\bar{z}) = 0$, because $\mathbb{P}[z_0 \leq \bar{z}] = 0$. Also, because $\epsilon \geq 0$ under F,

$$H(z) \leq \mathbb{P}_F[z_0 - \tau\epsilon \leq z|v > \bar{y}] = \mathbb{P}_F[y_1 \leq z|y_1 > \bar{y}] = G_{1|0}(z),$$

where the inequality used the fact that $\zeta(z;\eta) \equiv z - \eta\pi(z) \geq z$, and the second equality used $F \in \mathcal{F}^\star$. Using the law of total probability, H also satisfies

$$H_0(z) = H(z)\,\mathbb{P}[v > \bar{y}] + \mathbb{P}[z_0 \leq z|v \leq \bar{y}]\,\mathbb{P}[v \leq \bar{y}]$$

for all z. The derivative of the second term with respect to z is non-negative, so differentiating both sides with respect to z implies that

$$h_0(z) \geq H'(z)\,\mathbb{P}[v > \bar{y}] = H'(z)(1 - \bar{p}). \tag{54}$$

So, $H \in \mathcal{H}^\star$. Moreover, the definition of H ensures that

$$\int zH'(z)\,dz = \mathbb{E}_F[z_0|v > \bar{y}] = \mathbb{E}_F[z_0|y_1 > \bar{y}].$$

This establishes one direction of the set inclusion in (53).

Conversely, suppose that $H \in \mathcal{H}^\star$. Let u be a random variable that is independent of y_0 (so also z_0) and has a uniform distribution on $[0, 1]$. Let $\bar{u}(z) \equiv H'(z)/\bar{h}(z) \equiv H'(z)(1-\bar{p})/h_0(z)$. Note that because u and z_0 are independent, and u is uniformly distributed on $[0, 1]$,

$$\mathbb{P}[u > \bar{u}(z_0)] = 1 - \mathbb{E}[\bar{u}(z_0)] = 1 - \int \frac{H'(z)(1-\bar{p})}{h_0(z)}\, h_0(z)dz = \bar{p}, \tag{55}$$

because every $H \in \mathcal{H}^\star$ is a distribution function. Define the virtual earnings

$$v \equiv \begin{cases} q_{1|0}(H(z_0)), & \text{if } \mathbb{1}[u \le \bar{u}(z_0)]. \\ \bar{y} - \tau e, & \text{if } \mathbb{1}[u > \bar{u}(z_0)]. \end{cases} \tag{56}$$

where $q_{1|0}(p)$ denotes the pth quantile of y_1, conditional on $y_1 > \bar{y}$, and $e \ge 0$ is any non-negative constant. Use the virtual earnings to define the random variable

$$\tilde{\epsilon} \equiv \frac{1}{\tau}\left(z_0 - v\right), \tag{57}$$

and let F denote the distribution of $\tilde{\epsilon}$, conditional on y_0. We will show that $F \in \mathcal{F}^\star$, and that the conditional expectation of z_0 for the non-bunchers under F is equal to the expectation of H.

We start by showing that if $\tilde{\epsilon}$ follows distribution F, then the implied distribution of z_0 for the non-bunchers is just H. First, notice that

$$\begin{aligned} \mathbb{P}_F[z_0 \le z|y_1 > \bar{y}] &= \mathbb{P}_F[z_0 \le z|z_0 > \bar{y} + \tau\tilde{\epsilon}] \\ &= \mathbb{P}[z_0 \le z|v > \bar{y}] = \mathbb{P}[z_0 \le z|u \le \bar{u}(y_0)], \end{aligned} \tag{58}$$

where the first equality used (VE), the second used (57), and the final one used (56) together with $e \geq 0$. Manipulating (58) shows that

$$\begin{aligned}\mathbb{P}_F[z_0 \leq z | y_1 > \bar{y}] &= \mathbb{E}\left[\mathbb{1}[z_0 \leq z]\mathbb{1}[u \leq \bar{u}(z_0)]\right]/(1-\bar{p}) \\ &= \mathbb{E}\left[\mathbb{1}[z_0 \leq z]\,\mathbb{P}[u \leq \bar{u}(z_0)|z_0]\right]/(1-\bar{p}) \\ &= \mathbb{E}\left[\mathbb{1}[z_0 \leq z]\bar{u}(z_0)\right]/(1-\bar{p}) \\ &= \int_{-\infty}^{z} \frac{H'(w)}{\bar{h}(w)} \frac{h_0(w)}{1-\bar{p}}\, dw = H(z). \end{aligned} \tag{59}$$

An implication is that

$$\mathbb{E}_F[z_0 | y_1 > \bar{y}] = \int z H'(z)\, dz.$$

We complete the argument by showing that $F \in \mathcal{F}^\star$. First, $F \in \mathcal{F}^\dagger$ because

$$\begin{aligned}\mathbb{P}_F[\tilde{\epsilon} \geq 0] &= \mathbb{P}[v \leq z_0] \\ &= \mathbb{P}[v \leq z_0, u \leq \bar{u}(z_0)] + \mathbb{P}[v \leq z_0, u > \bar{u}(z_0)] \\ &= \mathbb{P}\left[q_{1|0}(H(z_0)) \leq z_0, u \leq \bar{u}(z_0)\right] + \mathbb{P}\left[\bar{y} - \tau e \leq z_0, u > \bar{u}(z_0)\right] \\ &= \mathbb{P}\left[H(z_0) \leq G_{1|0}(z_0), u \leq \bar{u}(z_0)\right] + \mathbb{P}\left[\bar{y} \leq z_0 + \tau e, u > \bar{u}(z_0)\right] = 1, \end{aligned}$$

where the final equality used $H \in \mathcal{H}^\star$ and the fact that $z_0 \equiv \zeta(y_0;\eta) \geq y_0 \geq \bar{y}$ together with $e \geq 0$. Second, to show that $F \in \mathcal{F}^\star$, we first use the definition of $\tilde{\epsilon}$ to see that $\mathbb{P}_F[y_1 \leq y] = \mathbb{P}_F[z_0 - \tau\tilde{\epsilon} \leq y] = \mathbb{P}[v \leq y]$. Then, we use the definition of v together with (55) to conclude that

$$\begin{aligned}\mathbb{P}_F[y_1 \leq y] &= \mathbb{P}[q_{1|0}(H(z_0)) \leq y | u \leq \bar{u}(z_0)](1-\bar{p}) \\ &\quad + \mathbb{P}[\bar{y} - \tau e \leq y | u > \bar{u}(z_0)]\bar{p} \\ &= \mathbb{P}[H(z_0) \leq G_{1|0}(y) | u \leq \bar{u}(z_0)](1-\bar{p}) + \bar{p} \\ &= G_{1|0}(y)(1-\bar{p}) + \bar{p} = G_1(y), \end{aligned}$$

where the second-to-last equality used the finding (58)–(59) that $\mathbb{P}_F[z_0 \leq z | u \leq \bar{u}(z_0)] = H(z)$, and the final equality used $\mathbb{P}[y_1 > \bar{y}] = 1 - \bar{p}$ together with $\mathbb{P}[y < \bar{y}] = 0$. This concludes the proof. Q.E.D.

Proposition SA.1 shows that we can establish sharpness by characterizing the smallest mean that can be formed from $H \in \mathcal{H}^\star$. The set $\mathcal{H}^\star$ depends on the observed distribution functions H_0 and $G_{1|0}$. Let h_0 and $g_{1|0}$ denote their respective densities, and recall the definition $\bar{h}(z) \equiv h_0(z)/(1-\bar{p})$. In what follows, we impose the following assumption, which says that $\bar{h}(z)$ and $g_{1|0}(z)$ only intersect finitely many times. We state the assumption in terms of their difference, which we denote as

$$\Delta(z) \equiv \frac{h_0(z)}{1-\bar{p}} - g_{1|0}(z).$$

Assumption SA.1. (Finite intersections) There exist a finite number of points $x_0 \equiv \bar{z} < x_1 < \cdots < x_{I-1} < x_I \equiv \infty$ such that the sign of $\Delta(z)$ is constant on each open interval (x_{i-1}, x_i), $i = 1, \ldots, I$.

We show in the following proposition that Assumption SA.1 makes it possible to explicitly characterize an $H \in \mathcal{H}^\star$ that attains the smallest mean and thereby also constructs γ_{LB}, the sharp lower bound for non-bunchers.

Proposition SA.2. Suppose that Assumption SA.1 holds. Then $\mathcal{H}^\star$ contains a function $H^\star$ with derivative given by

$$h^\star(z) \equiv \frac{d}{dz} H^\star(z) = \begin{cases} \frac{h_0(z)}{1-\bar{p}} & \text{if } H^\star(z) < G_{1|0}(z) \\ \min\left\{ g_{1|0}(z), \frac{h_0(z)}{1-\bar{p}} \right\} & \text{if } H^\star(z) \geq G_{1|0}(z) \end{cases} \tag{60}$$

for almost every z. Moreover, $H^\star(z) \geq H(z)$ for any other $H \in \mathcal{H}^\star$ and all $z \geq \bar{z}$. As a consequence,

$$\gamma_{\text{LB}} = \frac{1}{\tau}\left(\int z h^\star(z)\, dz - \mathbb{E}[y_1 | y_1 > \bar{y}] \right).$$

Proof of Proposition SA.2. We proceed in three steps. First, we show that the claimed function $H^\star$ exists. Second, we show that $H^\star \in \mathcal{H}^\star$. Third, we show that $H^\star(z) \geq H(z)$ for any other $H \in \mathcal{H}^\star$ and all $z \geq \bar{z}$. This third point implies that $\int z h^\star(z)\, dz \leq \int z dH(z)$ for any other $H \in \mathcal{H}^\star$. Appealing to Proposition SA.1 then provides the sharpness conclusion.

1. Existence. We construct $H^\star$ recursively on the intervals $(x_{i-1}, x_i]$ defined by Assumption SA.1. Set $H^\star(\bar{z}) = 0$, noting that this is necessary for $H^\star \in \mathcal{H}^\star \subseteq \mathcal{H}$ and that it trivially implies $H^\star(\bar{z}) \leq G_{1|0}(\bar{z})$. Suppose that $H^\star$ has been constructed on $[\bar{z}, x_{i-1}]$ for some i, with $H^\star(z) \leq G_{1|0}(z)$ for all $z \in [\bar{z}, x_{i-1}]$, and derivative satisfying (60). We will show how to extend the construction to $(x_{i-1}, x_i]$ while ensuring that $H^\star(z) \leq G_{1|0}(z)$ for all $z \in (x_{i-1}, x_i]$.

By Assumption SA.1, $\Delta(z)$ has constant sign on (x_{i-1}, x_i). This creates two cases which we treat separately.

The first case is when the sign of $\Delta(z)$ is negative, so that $h_0(z)/(1-\bar{p}) < g_{1|0}(z)$ for all $z \in (x_{i-1}, x_i)$. Define $H^\star(z)$ for $z \in (x_{i-1}, x_i]$ as

$$H^\star(z) \equiv H^\star(x_{i-1}) + \int_{x_{i-1}}^{z} \frac{h_0(x)}{1-\bar{p}}\, dx. \tag{61}$$

Because the integrand is integrable, $H^\star$ is absolutely continuous on $(x_{i-1}, x_i]$. Because $H^\star(x_{i-1}) \leq G_{1|0}(x_{i-1})$ by the induction assumption and $h_0(z)/(1-\bar{p}) < g_{1|0}(z)$, we conclude that $H^\star(z) < G_{1|0}(z)$ for all $z \in (x_{i-1}, x_i]$. Differentiating (61) shows that (60) holds for almost every $z \in (x_{i-1}, x_i]$.

The second case is when the sign of $\Delta(z)$ is positive, so that $h_0(z)/(1-\bar{p}) > g_{1|0}(z)$ for all $z \in (x_{i-1}, x_i)$. We make use of an intermediate construction involving the function defined as

$$\tilde{H}(z) \equiv H^\star(x_{i-1}) + \int_{x_{i-1}}^{z} \frac{h_0(s)}{1-\bar{p}} ds. \quad \text{for all } z \in (x_{i-1}, x_i].$$

The induction assumption is that $\tilde{H}(x_{i-1}) = H^\star(x_{i-1}) \leq G_{1|0}(x_{i-1})$. Let t_i denote the first point at which $\tilde{H}(z)$ crosses $G_{1|0}(x_{i-1})$ on $[x_{i-1}, x_i]$:

$$t_i \equiv \inf\{z \in [x_{i-1}, x_i] : \tilde{H}(z) \geq G_{1|0}(z)\}.$$

If the set is empty, then we let $t_i = x_i$. We conclude that $\tilde{H}(z) \leq G_{1|0}(z)$ for all $z \in (x_{i-1}, t_i]$ and $\tilde{H}(z) \geq G_{1|0}(z)$ for all $z \in [t_i, x_i]$.

Now we use $\tilde{H}$ to define $H^\star$ on $(x_{i-1}, x_i]$ as

$$H^\star(z) \equiv \begin{cases} \tilde{H}(z) & z \in (x_{i-1}, t_i), \\ \tilde{H}(t_i) + \int_{t_i}^{z} g_{1|0}(s) ds & z \in [t_i, x_i]. \end{cases}$$

Then $H^\star(z) \leq G_{1|0}(z)$ for all $z \in (x_{i-1}, t_i)$ and $H^\star(z) = G_{1|0}(z)$ for all $z \in (t_i, x_i]$. It follows immediately from the definition of $H^\star$ that it is absolutely continuous. When $H^\star(z) < G_{1|0}(z)$, so that $z < t_i$, the derivative of $H^\star$ is $h^\star(z) = \tilde{h}(z) = h_0(z)/(1-\bar{p})$ for $z < t_i$, which satisfies (60). When $H^\star(z) = G_{1|0}(z)$, so that $z \geq t_i$, the derivative of $H^\star$ is $g_{1|0}(z)$, which is smaller than $h_0(z)/(1-\bar{p})$ for all $z \in (x_{i-1}, x_i]$ because $\Delta(z) \geq 0$. We conclude that (60) is satisfied for all $z \in (x_{i-1}, x_i]$.

2. $H^\star \in \mathcal{H}^\star$. We now show that $H^\star \in \mathcal{H}^\star$. We have already established that $H^\star$ is absolutely continuous and strictly increasing with $h^\star(z) \leq h_0(z)/(1-\bar{p})$ and that $H^\star(z) \leq G_{1|0}(z)$ for all $z \geq \bar{z}$. It remains to show that $H^\star$ is a well-defined distribution on $[\bar{z}, \infty)$. Because $H^\star(\bar{z}) = 0$, the only remaining property we need to verify is that $\lim_{z\to\infty} H^\star(z) = 1$.

We have shown that our construction of $H^\star$ either satisfies $H^\star(z) = G_{1|0}(z)$ or $H^\star(z) < G_{1|0}(z)$ and that it switches between these cases at only a finite number of $z \geq \bar{z}$ determined by the points t_i. This implies that there exists a terminal $t^\dagger \geq \bar{z}$ such that either $H^\star(z) = G_{1|0}(z)$ for all $z \geq t^\dagger$ or $H^\star(z) < G_{1|0}(z)$ for all $z \geq t^\dagger$. The first case immediately implies the desired property that $\lim_{z\to\infty} H^\star(z) = \lim_{z\to\infty} G_{1|0}(z) = 1$, because $G_{1|0}(z)$ is itself a distribution function.

Consider the second case, in which $H^\star(z) < G_{1|0}(z)$ for all $z > t^\dagger$. Then $\lim_{z\to\infty} H^\star(z) \leq 1$. Because $H^\star$ and $G_{1|0}$ cross only a finite number of times, we know that $H^\star(t^\dagger) = G_{1|0}(t^\dagger)$. For $z > t^\dagger$ we get from (60) that

$$H^\star(z) = H^\star(t^\dagger) + \int_{t^\dagger}^{z} \frac{h_0(s)}{1-\bar{p}} ds = G_{1|0}(t^\dagger) + \frac{H_0(z) - H_0(t^\dagger)}{1-\bar{p}}.$$

Taking limits gives

$$\begin{aligned}
\lim_{z\to\infty} H^\star(z) &= G_{1|0}(t^\dagger) + \frac{1 - H_0(t^\dagger)}{1-\bar{p}} \\
&\geq G_{1|0}(t^\dagger) + \frac{1 - G_1(t^\dagger)}{1-\bar{p}} \\
&= G_{1|0}(t^\dagger) + \frac{1 - \bar{p} - G_{1|0}(t^\dagger)(1-\bar{p})}{1-\bar{p}} = 1,
\end{aligned}$$

where the inequality used Proposition 4 together with our premise that the model is not misspecified. We conclude that $\lim_{z\to\infty} H^\star(z) = 1$, so that $H^\star \in \mathcal{H}^\star$.

3. Maximality. Consider some $H \in \mathcal{H}^\star$. Suppose by way of contradiction that $H(z) > H^\star(z)$ for some $z > \bar{z}$. Define

$$\tilde{z} \equiv \inf\{z > \bar{z} : H(z) > H^\star(z)\}.$$

Then $H(z) \leq H^\star(z)$ for all $z < \tilde{z}$. Because $\mathcal{H}^\star$ contains only continuous functions, $H(\tilde{z}) = H^\star(\tilde{z})$ and there exists a $\delta > 0$ such that $H(z) > H^\star(z)$ for all $z \in (\tilde{z}, \tilde{z} + \delta)$.

Let $D(z) \equiv H(z) - H^\star(z)$. Then D is absolutely continuous, $D(\tilde{z}) = 0$, and $D(z) > 0$ on $(\tilde{z}, \tilde{z} + \delta)$. This implies that D has a derivative $D'(z) = h(z) - h^\star(z)$ that is strictly positive on a set of positive measure $\mathcal{Z}$ contained in $(\tilde{z}, \tilde{z} + \delta)$. Let $s \in \mathcal{Z}$ be any point at which H, $H^\star$, and D are all differentiable, so that $h(s) > h^\star(s)$, noting that the complement set has measure zero. Then $G_{1|0}(s) \geq H(s) > H^\star(s)$. However, the definition of $H^\star$—in particular (60)—implies that

$$h(s) > h^\star(s) = \frac{h_0(s)}{1-\bar{p}},$$

which contradicts the properties of $H \in \mathcal{H}^\star$. We conclude that no such H can exist.

Q.E.D.

SA.5 Computational approach

In this appendix, we develop a computational method for constructing bounds under the assumption that (ϵ, η) has known finite support.

Condition FS. (Finite support) $\mathcal{F}^\dagger$ is the subset of $\mathcal{F}$ under which $\sum_{j=1}^{J} \mathbb{P}_F[\epsilon = \epsilon_j, \eta = \eta_j] = 1$, where $\{(\epsilon_j, \eta_j)\}_{j=1}^{J} \subset [0,\infty) \times [-1, 0]$ is a known set.

Consider any $F \in \mathcal{F}^\star$, where $\mathcal{F}^\star$ is given by (ID). Let $\{\mathcal{Y}_{0,k}\}_{k=1}^{K}$ be a partition of the support of y_0. If $\bar{\bar{y}}_0 \neq \infty$, then make the last interval $\mathcal{Y}_{0,K} = (\bar{\bar{y}}_0, \infty)$. Because $F \in \mathcal{F}^\star$, $\mathbb{P}[v_0 \in \mathcal{Y}_{0,k}] = \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]$ is known for all k. Then define

$$\pi_{F(j|k)} \equiv \mathbb{P}_F[\epsilon = \epsilon_j, \eta = \eta_j | v_0 \in \mathcal{Y}_{0,k}].$$

Let π_F collect the shares $\pi_{F(j|k)}$ over all (j,k).

SA.5.1 Target parameters

We consider any parameter of the form $c'\pi_F$ for known c. Examples include averages of the labor supply elasticities such as

$$\mathbb{E}_F[\epsilon^u] = \sum_{j,k} \pi_{F(j|k)} \times (\epsilon_j + \eta_j)\, \mathbb{P}_F[v_0 \in \mathcal{Y}_{0,k}] = \sum_{j,k} \pi_{F(j|k)} \times \underbrace{(\epsilon_j + \eta_j)\, \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]}_{\text{known}},$$

where the second equality holds for any $F \in \mathcal{F}^\star$. Conditional averages among subpopulations defined by partition points can also be expressed in this form. For example, if $\mathcal{Y}_{0,k^\star}$ has right endpoint $y_0^\star$, then

$$\mathbb{E}_F[\epsilon^u | v_0 \leq y_0^\star] = \sum_{j,k} \pi_{F(j|k)} \times \underbrace{\frac{(\epsilon_j + \eta_j)\mathbb{1}[k < k^\star]\, \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]}{\mathbb{P}[y_0 \leq y_{0,k^\star}]}}_{\text{known}}.$$

SA.5.2 The identified set

If $F \in \mathcal{F}^\star$, then the possible values of π_F are constrained by the observed distribution of y_1. For any $y \geq \bar{y}$, we must have

$$G_1(y) = \sum_{j,k} \pi_{F(j|k)} \, \mathbb{P}_F[\nu(v_0, \epsilon_j, \eta_j) \leq y | \epsilon = \epsilon_j, \eta = \eta_j, v_0 \in \mathcal{Y}_{0,k}] \, \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}], \tag{62}$$

where we again used that $\mathbb{P}_F[v_0 \in \mathcal{Y}_{0,k}] = \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]$.

If $\mathcal{Y}_{0,k}$ were a singleton, then the conditional probability in (62) would be either zero or one. Because $\mathcal{Y}_{0,k}$ is an interval, this probability may be unknown for some (j,k). Because $\nu(v_0, \epsilon, \eta)$ is strictly increasing in v_0, lower and upper bounds on these probabilities are obtained by evaluating ν at the interval endpoints:

$$\mathcal{I}_{\mathrm{L}}(y) \equiv \{(j,k) : \nu(y_{0,k+1}, \epsilon_j, \eta_j) \leq y\} \quad \text{and} \quad \mathcal{I}_{\mathrm{U}}(y) \equiv \{(j,k) : \nu(y_{0,k}, \epsilon_j, \eta_j) \leq y\}.$$

It follows that

$$G_1(y) \geq \sum_{j,k} \pi_{F(j|k)} \underbrace{\mathbb{1}[(j,k) \in \mathcal{I}_{\mathrm{L}}(y)] \, \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]}_{\text{known}}, \tag{63}$$

$$\text{and} \quad G_1(y) \leq \sum_{j,k} \pi_{F(j|k)} \underbrace{\mathbb{1}[(j,k) \in \mathcal{I}_{\mathrm{U}}(y)] \, \mathbb{P}[y_0 \in \mathcal{Y}_{0,k}]}_{\text{known}}. \tag{64}$$

Define the set that satisfies (63)–(64) as

$$\Pi^\star \equiv \Big\{ \pi_F \; : \pi_{F(j|k)} \in [0,1] \; \forall \; (j,k), \; \sum_j \pi_{F(j|k)} = 1 \; \forall \; k = 1, \ldots, K, \\ \text{(63)-(64) hold} \; \forall \; y \in \{y_{1,m}\}_{m=1}^M \Big\}.$$

Then $\Pi^\star$ is an outer set in the sense that $F \in \mathcal{F}^\star$ implies that $\pi_F \in \Pi^\star$. Outer bounds on the target parameters can be computed by optimizing over $\Pi^\star$:

$$\beta^\star_{\mathrm{L}} \equiv \min_{\pi \in \Pi^\star} c'\pi \quad \text{and} \quad \beta^\star_{\mathrm{U}} \equiv \max_{\pi \in \Pi^\star} c'\pi. \tag{65}$$

These optimization programs are linear programs with finitely many constraints. Their optimal values constitute outer bounds: if $F \in \mathcal{F}^\star$, then $c'\pi_F \in [\beta^\star_{\mathrm{L}}, \beta^\star_{\mathrm{U}}]$.

SA.5.3 Imposing additional assumptions

Additional assumptions can be imposed by incorporating additional linear constraints on π_F into the definition of $\Pi^\star$. An example of such an assumption is that η is a known and deterministic constant, which can be imposed by setting $\pi_{F(j|k)} = 0$ whenever $\eta_j \neq \eta$. Another example is the assumption that $\mathbb{E}[\epsilon | y_0 \in \mathcal{Y}_{0,k}]$ is bounded for all k.

We can also impose assumptions on buncher and non-buncher averages by replacing $\pi_{F(j|k)}$ with $\pi_{F(j,b|k)} \equiv \mathbb{P}[\epsilon = \epsilon_j, \eta = \eta_j, B = b \mid v_0 \in \mathcal{Y}_{0,k}]$ for $b = 0, 1$. We restrict attention to primitives that satisfy the restrictions implied by bunching: set $\pi_{F(j,0|k)} = 0$ whenever no $v_0 \in \mathcal{Y}_{0,k}$ satisfies $\nu(v_0, \epsilon_j, \eta_j) > \bar{y}$ and set $\pi_{F(j,1|k)} = 0$ whenever no $v_0 \in \mathcal{Y}_{0,k}$ satisfies $\nu(v_0, \epsilon_j, \eta_j) \leq \bar{y}$. Apart from this refinement, the same approach applies.

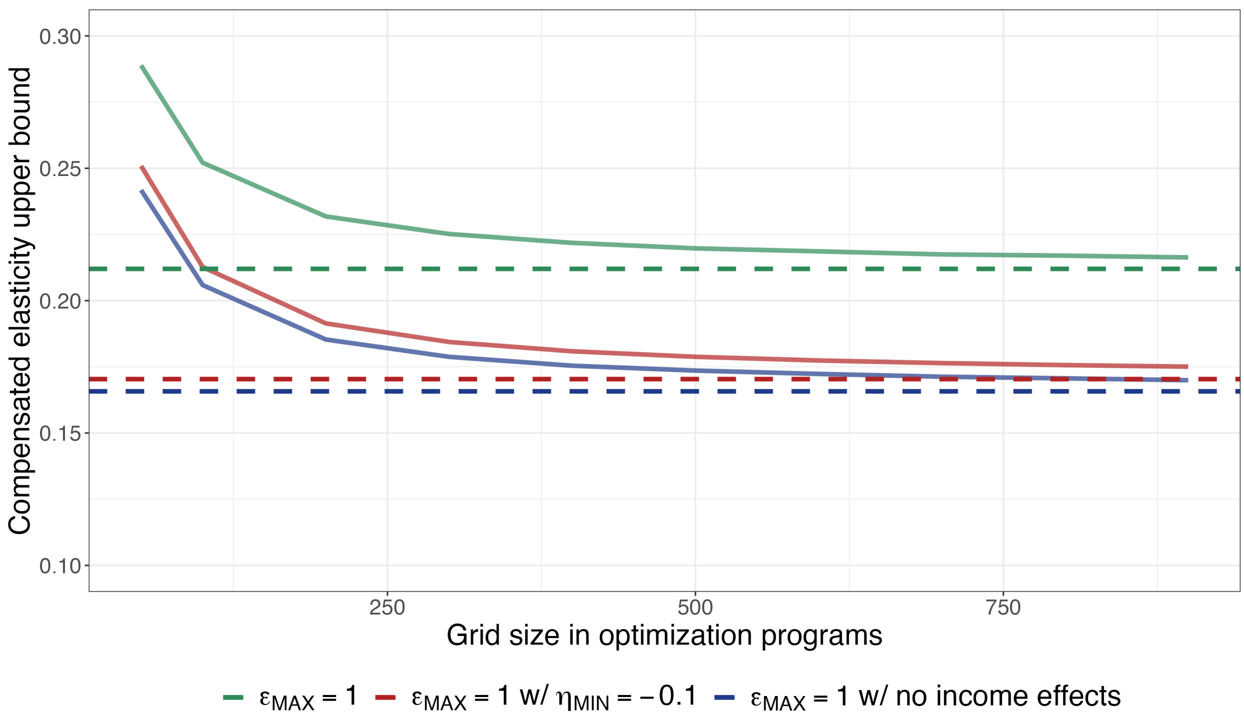
SA.5.4 Empirical results

We apply the computational approach to compute approximate sharp bounds for our results in Section 8. We discretize ϵ and η into 50 and 30 values, respectively, concentrating mass near zero and allowing the compensated elasticity to have a long upper tail. We have experimented with different discretization strategies and found similar conclusions.

As in Section 8, we use the distribution of y_0 up to the 10th quantile and the full distribution of y_1. We partition the observed portions of y_0 and y_1 into quantile-spaced intervals. We construct bounds on $\mathbb{E}[\epsilon \mid \bar{y} < v_0 \leq y_0^\star]$ and $\mathbb{E}[\epsilon^u \mid \bar{y} < v_0 \leq y_0^\star]$, where $y_0^\star$ is the 10th quantile. These are the same parameters reported in Table 4. To assess sensitivity to discretization, we vary the number of intervals in each partition.

We first compare the computational upper bounds to the sharp upper bounds in Table 4. For coarse partitions, the computational upper bounds lie noticeably above the sharp analytic bounds. As the grid is refined, they quickly converge. Figure SA.5 shows this pattern for the compensated elasticity under three assumptions: (i) $\epsilon_{\text{MAX}} = 1$ in green, (ii) $\epsilon_{\text{MAX}} = 1$ and $\eta_{\text{MIN}} = -0.1$ in red, and (iii) $\epsilon_{\text{MAX}} = 1$ and no income effects in blue. The dotted lines show the sharp upper bounds from Table 4, and the solid curves show the computational upper bounds as the grid size varies. With 500 intervals, the computational upper bounds are within 0.01 of the sharp upper bounds. The same pattern holds for the upper bounds on the uncompensated elasticities, which do not vary across the maintained assumptions in Table 4. The computational lower bounds are always zero, regardless of the grid, reinforcing our conclusion that zero is indeed the sharp lower bound.

Figure SA.5: Comparing explicit and computational bounds



Notes: This figure presents upper bounds on $\mathbb{E}[\epsilon|y_0 \leq y_0^\star]$ *under three different assumptions:* $\epsilon_{\text{MAX}} = 1$ *in green,* $\epsilon_{\text{MAX}} = 1$ *and* $\eta_{\text{MIN}} = -0.1$ *in red, and* $\epsilon_{\text{MAX}} = 1$ *and no income effects in blue. The dotted lines depict the sharp upper bounds from Table 4, and the solid curves depict the upper bounds obtained from the computational approach as we vary the grid size.*