Substitution and income effects of labor income taxation

Michael Graber [*] Morten Håvarstein[†] Magne Mogstad [‡]
Gaute Torsvik [§] Ola L. Vestad [¶]

December 14, 2025

Latest version available here.

Abstract

The elasticity of taxable income (ETI) parameter is a key quantity in empirical analysis of tax policy and labor supply. We examine when a commonly applied class of ETI estimands can be used to learn about individuals' ETI parameters and their (un)compensated elasticities of labor supply. We begin by providing necessary and sufficient conditions for these estimands to be given a causal interpretation as a positively weighted average of heterogeneous ETI parameters. We then apply these results to empirically analyze a reform of the Norwegian tax system that reduced the marginal tax rates on middle and high incomes. The estimated ETI parameters increase steadily with income, meaning high-income individuals are more responsive to tax changes than middle-income individuals. Next, we show how (un)compensated elasticities of labor supply can be bounded directly from the ETI estimands, or point identified by combining these estimands with estimates of earnings responses to lottery winnings. The results suggest an (un)compensated elasticity of 0.15 (0.0) for middle-income individuals. The (un)compensated elasticity estimates increase steadily with income to around 0.5 (0.35) for high-income individuals. These findings imply a substantial excess burden of taxation, and that reducing the top-income tax rate would increase tax revenue. Under separable utility, our findings are also informative about how the intertemporal elasticity of substitution and the Frisch elasticity vary across the income distribution.

[*]Statistics Norway.

[†]Kenneth C. Griffin Department of Economics, University of Chicago; Department of Economics, University of Oslo.

[‡]Kenneth C. Griffin Department of Economics, University of Chicago; Statistics Norway; NBER.

[§]Department of Economics, University of Oslo.

[¶]Statistics Norway.

Håvarstein, Mogstad, and Torsvik appreciate the financial support from the Norwegian Research Council and NSF Grant 341250. We thank Deniz Dutz and Harald Uhlig for valuable comments.

1 Introduction

The elasticity of taxable income (ETI) parameter measures how taxable income responds to reforms that change the marginal tax rate. It is a key quantity in tax policy for assessing how exogenous changes in tax rates will causally affect income and tax revenue (Auten and Carroll, 1999). Also, it is often interpreted as a compensated elasticity of labor supply (Saez et al., 2012), which can be used to assess the excess burden of taxation (Feldstein, 1995). The goal of our paper is to show when and how a commonly applied class of ETI estimands can be used to learn about individuals' ETI parameters and their (un)compensated elasticities of labor supply.

In Section 2, we provide necessary and sufficient conditions for the ETI estimands to be given a causal interpretation as a positively weighted average of heterogeneous ETI parameters. This identification result is constructive, leading to empirical specifications and estimators that can be easily implemented. These specifications differ from the ETI estimands commonly used, which fail to satisfy the conditions for a causal interpretation when elasticities are heterogeneous. A causal interpretation with heterogeneous elasticities requires both a "saturated" specification that controls for past income non-parametrically, and no comparison of earnings responses between individuals who experience changes in marginal tax rates of varying degrees.

The identification results in Section 2 guide our empirical analysis in Section 3 of a reform of the Norwegian tax system that reduced the marginal tax rates on middle and high incomes.[1] We find that commonly used ETI estimands that are causal only under constant elasticities understate the average ETI parameter. The specifications that can be given a causal interpretation produce estimates of an average ETI parameter across the income distribution of around 0.23. This average misses a great deal: the estimated ETI parameters increase steadily with income, meaning high-income individuals are more responsive to tax changes than middle-income individuals. For example, the ETI parameter at the median income is less than 0.1, while it is 0.35 at the 90th percentile.

In Section 4, we show how (un)compensated elasticities of labor supply can be point or partially identified from our preferred causal specification of the ETI estimands. The continuous choice labor supply model we consider allows for income effects, elasticities that vary across the income distribution, and correlation between individuals' productivity and their unobserved taste for work.[2] This model allows us to express the ETI estimands in terms

[1] This reform has been previously analyzed by Vattø (2020), who estimates an average ETI of 0.15 (see also Thoresen and Vattø, 2015 and Berg and Thoresen, 2020).

[2] Our model nests typical labor supply models in public finance (e.g., Feldstein, 1999; Gruber and Saez, 2002 and Saez et al., 2012). An alternative class of models represents the budget with a discrete set of alternatives and assumes that individuals' productivity is independent of their unobserved taste for work. This taste also follows a known parametric distribution. See Aaberge and Colombino (2018) and the references therein for discussion of these models and the findings they produce.

of the (un)compensated elasticities plus a bias term that would be observable in the data if income effects were known or could be estimated. This observation motivates and guides the analysis in the remainder of the paper, where we consider different approaches to point or partially identify the labor supply elasticities and functionals of these elasticities, such as the excess burden of taxation.

The first approach we consider is to construct bounds using the Engel aggregation condition, which implies that income effects are bounded between -1 and 0. The bounds suggest that the compensated and uncompensated elasticities of high-income individuals at the 90th percentile are at least 0.35 and 0.1, respectively. By comparison, the compensated elasticity of middle-income individuals close to the median income is bounded between 0.05 and 0.5, while their uncompensated elasticities are close to zero or negative. These bounds can be tightened considerably by ruling out implausibly small and large income effects. Both the upper bound on the compensated elasticities and the lower bound on the uncompensated elasticities then become more informative.

There are two ways to move from partial to point identification. One is to make stronger assumptions. Gruber and Saez (2002) assumes that income and substitution effects do not vary across the income distribution. The (un)compensated elasticity is then point identified. We test this assumption and find that it is strongly rejected by the data.

Another possible way to achieve point identification, which we consider in Section 5, is to combine the ETI estimands with external estimates of income effects. A number of studies have estimated income effects from lottery winnings that create plausibly exogenous variation in unearned income, holding fixed all other determinants of behavior, such as preferences and wages.[3] We follow this approach and use Norwegian data on lottery winnings to estimate earnings and employment responses to exogenous changes in unearned income. We show how these responses allow us to infer the income effects on the intensive margin that we need to point identify the (un)compensated elasticities. The resulting point estimates suggest an (un)compensated elasticity of 0.1 (-0.05) for middle-income individuals at the median income. The (un)compensated elasticity estimates increase steadily with income to around 0.45 (0.3) for high-income individuals at the 90th percentile.

In Section 6, we show that the findings in Section 5 imply a substantial excess burden of taxation, and that reducing the top-income tax rate would significantly increase tax revenue. We find that these conclusions contrast sharply with the results we obtain from conventional calculations that use the common ETI estimand, ignore income effects, and assume constant labor supply elasticities. The excess burden then becomes much lower, and the revenue-maximizing top-income tax rate moves closer to the observed one.

In Section 7, we consider the implications of our findings for the elasticity of intertemporal substitution (EIS) of consumption and the Frisch elasticity. The EIS determines how

[3] See, for example, Cesarini et al. (2017), Golosov et al. (2024), and Imbens et al. (2001).

agents adjust consumption in response to intertemporal price changes, while the Frisch elasticity measures the response of labor supply to predictable wage changes. We assume the same parametric specification of preferences as in Heathcote et al. (2014). Given this specification, compensated elasticities and income effects map directly to EIS and Frisch elasticities. The resulting Frisch elasticity (EIS) is 0.1 (0.5) for middle-income individuals at the median income. The Frisch elasticity (EIS) estimates increase steadily with income to around 0.5 (2) for high-income individuals at the 90th percentile. We show how the heterogeneity in these elasticities across the income distribution matters for conclusions about the efficiency loss of temporary tax changes.

Our paper contributes to a large set of studies that have used models of labor supply to try to recover income and substitution effects from observational variation in unearned income, wages, and tax rates. The models, data, and findings have been summarized and critiqued in multiple review articles, including Blundell and Macurdy (1999), Keane (2011), Killingsworth and Heckman (1986), Pencavel (1986), and Saez et al. (2012). As emphasized in these reviews, there is no consensus about the size of income and substitution effects and how they vary across the income distribution. As a result, it is difficult to draw credible conclusions about parameters that depend on these income and substitution effects, such as the excess burden of taxation and the revenue-maximizing tax rates.

A key reason for the lack of consensus is that it has been difficult to separately identify income and substitution effects without strong assumptions on functional form and the distribution of unobservables. Our paper shows how to construct informative bounds on (un)compensated elasticities from variation in take-home pay that arises from tax reforms. We also show how income and substitution effects can be point identified by combining the variation in take-home pay that arises from tax reforms with plausibly exogenous changes in unearned income. By doing so, our paper offers credible evidence on income and substitution effects, how they vary across the income distribution, and their implications for the excess burden of taxation and the revenue-maximizing top-income tax rate.

Our paper also contributes to the ETI literature, which analyzes how taxable income responds to reforms that change the marginal tax rates.[4] We extend the usual identification argument to allow for heterogeneity in the individuals' ETI parameters. Our necessary and sufficient conditions offer a blueprint for estimating ETIs while allowing for heterogeneous elasticities. The variation in elasticities across the income distribution that we find is significant, both statistically and economically. An average elasticity is therefore far from sufficient to predict how counterfactual changes in marginal tax rates would affect aggregate earnings and tax revenue.

[4]Notable contributions to this literature include Auten and Carroll (1999), Burns and Ziliak (2017), Feldstein (1995), Gruber and Saez (2002), and Kleven and Schultz (2014). See Saez et al. (2012) for a review and Neisser (2021) for a meta-analysis.

A related paper is Kumar and Liang (2020), which considers the causal interpretation of the ETI estimand without covariates, under the assumption that the tax system is randomly assigned across individuals. They show that this estimand is generally not equal to a *particular* weighted average of individual ETI parameters, unless the instrument is valid and the individual ETI parameters are homogeneous.

Our results about the ETI estimand differ in several ways. First, we consider the causal interpretation of the ETI estimand with covariates and do not assume that the tax system is randomly assigned. The inclusion of covariates in the theoretical results is important, since empirical work tries to flexibly control for individual characteristics such as past income. Second, we provide sufficient and necessary conditions for the ETI estimand to recover *any* positively weighted average of individual ETI parameters. This is arguably a minimal requirement for the ETI estimand to be an interesting quantity, but it is not sufficient. We therefore strengthen our result by showing *the* causal interpretation of the ETI estimand as a specific positively weighted average of individual ETI parameters. Third, we consider the bias of the ETI estimand due to income effects and how it can be corrected for by constructing bounds, invoking auxiliary assumptions, or using additional data.

Our paper is also related to a set of empirical studies that have used lottery winnings to estimate wealth and income effects (Bulman et al., 2021; Cesarini et al., 2017; Golosov et al., 2024; Imbens et al., 2001 and Picchio et al., 2018). To measure how lottery winnings are allocated over time, most of these studies rely on either the capitalization or the annuitization approach. In contrast, our rich administrative data allow for imputing consumption and savings over time (Eika et al., 2020), which means we can measure unearned income in each period without relying on additional assumptions. Thus, our paper offers credible evidence on income effects and how they vary across the income distribution without relying on assumptions about how households allocate their wealth over time.

Lastly, our paper provides new evidence on the Frisch elasticity and the EIS, and how these elasticities vary across the income distribution. The Frisch elasticity plays a key role in the design of tax policy (see e.g., Conesa et al., 2009) and in business cycle analysis (see e.g. Prescott, 2006). The EIS is relevant both for understanding consumption and saving behavior (Hall, 1988) and for the design of long-run capital taxation (Straub and Werning, 2020). Because of their importance, a large body of work attempts to estimate the Frisch elasticity (see e.g., MaCurdy, 1981 and Altonji, 1986) and the EIS (see e.g., Hansen and Singleton, 1983 and Attanasio and Weber, 1995). Keane and Neal (2025) review the literature on Frisch elasticities, arguing that estimates vary considerably across studies. Havránek (2015) reviews the EIS literature, concluding that estimates vary greatly between studies. An important question is what explains the differences in findings. A number of possible explanations have been discussed (see, e.g., Attanasio et al., 2018; Browning and Lusardi, 1996 and Chetty et al., 2011). We contribute by showing that both the EIS and the Frisch elasticity vary considerably due to population heterogeneity in the earnings responses at the intensive margin. The elasticities are small but significant for middle-income individuals, and fairly large and significant for individuals with higher incomes.

2 Interpretation of ETI estimands

We now introduce and analyze a class of commonly used ETI estimands. We first describe the research design and data that form the basis for these estimands. We then discuss the causal interpretation of these estimands, providing necessary and sufficient conditions for causality.

2.1 Research design

Figure 1 describes the research design and the data that it will use. The figure illustrates that the tax system changed from T_0 to T_1 in 2006, which affected the marginal tax rates on incomes above $\bar{Y}$. The reform cohorts ($G = 1$), which consist of observations from 2003-2007, experienced the *actual* reform in 2006. The placebo cohorts ($G = 0$), consisting of observations from 1998-2002, experience no change in the tax system. For each cohort type, we divide the data into a pre-period consisting of the first three years (event time $t = -3, -2, -1$) and a post-period covering the last two years ($t = 0, 1$).

We observe the earnings Y and marginal tax rates τ for every individual at each event time:

$$\tau_t = T_0'(Y_t) + \mathbb{1}[t \geq 0]G\Big(T_1'(Y_t) - T_0'(Y_t)\Big),$$

where T_d' denotes the derivative of tax function T_d. Thus, the marginal tax rate of an individual depends on her earnings Y and whether she faces the old (T_0) or the reformed tax system (T_1).

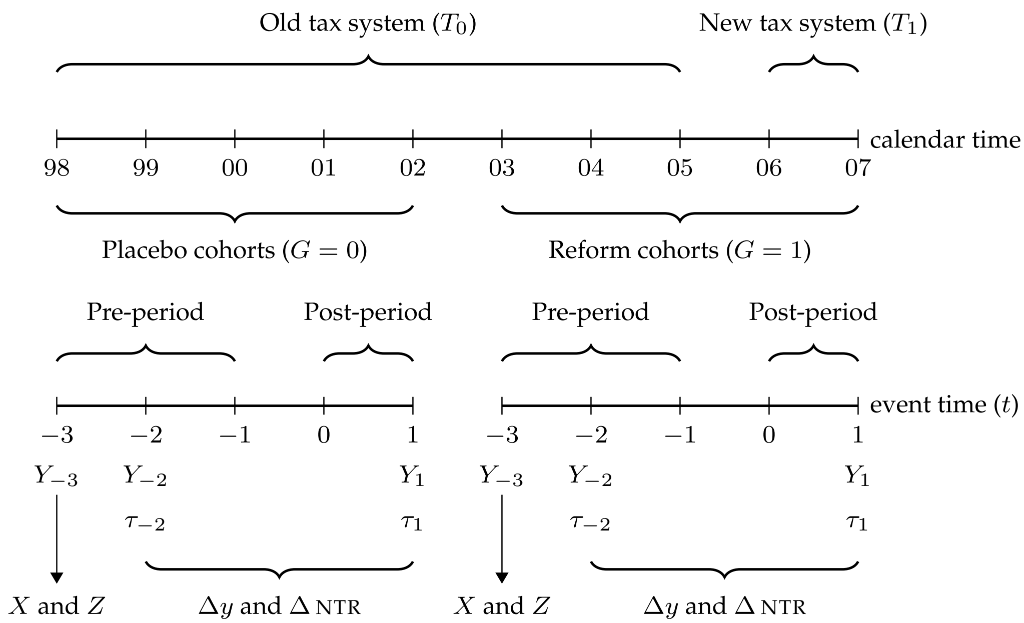
To address the simultaneity between marginal tax rates and earnings, the empirical ETI literature uses *simulated instruments*, defined as the predicted percentage change in net-of-tax rates because of the reform:

$$Z \equiv \log\left(\frac{1 - T_1'(X)}{1 - T_0'(X)}\right), \tag{1}$$

where X is earnings at event time -3.

The variable Z can take more than two values. It not only captures that the reform changes the marginal tax rates of some (treatment group with $Z \neq 0$), but not all individuals (control group with $Z = 0$), but also that the magnitude of the change may differ across earnings levels. Hence, treatment intensity can vary across treated individuals.

Figure 1: Anatomy of the research design



Notes: The figure presents the research design, illustrates the timing of the reform, and introduces notation.

2.2 ETI estimands and regression model

A possible estimand for the elasticity of taxable income (ETI) parameter is the difference-in-differences estimand that compares the earnings- and marginal tax rate growth of the treatment group to that of the control group:

$$\beta^{\text{DD}} \equiv \frac{\overbrace{\mathbb{E}[\Delta y \mid G=1, Z \neq 0] - \mathbb{E}[\Delta y \mid G=1, Z=0]}^{\equiv RF \text{ (Earnings DiD)}}}{\underbrace{\mathbb{E}[\Delta\,\text{NTR} \mid G=1, Z \neq 0] - \mathbb{E}[\Delta\,\text{NTR} \mid G=1, Z=0]}_{\equiv FS \text{ (Net-of-tax rate DiD)}}},$$

where $\Delta y \equiv \log Y_1 - \log Y_{-2}$, and $\Delta\,\text{NTR} \equiv \log(1-\tau_1) - \log(1-\tau_{-2})$.

A concern with this estimand is that because the reform changed marginal tax rates on incomes above $\bar{Y}$, the treated group has higher initial incomes than the control group. Thus, if individual income growth depends on initial income, either due to mean reversion or differential underlying income growth, parallel trends are unlikely to hold (Weber, 2014). To address these concerns, the literature uses the placebo cohorts to estimate the difference in earnings and net-of-tax rate growth between the treatment and control groups over the period when no tax reform occurred. The triple difference estimand β^{DDD} subtracts these placebo differences from the reform ones:

$$\beta^{\text{DDD}} \equiv \frac{\overbrace{RF}^{\text{Earnings DiD}} - \overbrace{\big(\mathbb{E}[\Delta y \mid G=0,\, Z\neq 0] - \mathbb{E}[\Delta y \mid G=0,\, Z=0]\big)}^{\text{Placebo earnings DiD}}}{\underbrace{FS}_{\text{Net-of-tax rate DiD}} - \underbrace{\big(\mathbb{E}[\Delta\,\text{NTR} \mid G=0,\, Z\neq 0] - \mathbb{E}[\Delta\,\text{NTR} \mid G=0,\, Z=0]\big)}_{\text{Placebo net-of-tax rate DiD}}}. \tag{2}$$

The triple difference estimand in (2) is nested by the following two-stage least squares (TSLS) regression model:

$$\Delta y = \alpha_0^y G + \beta\Delta\,\text{NTR} + f(X;\alpha^y) + u^y, \tag{3}$$

$$\Delta\,\text{NTR} = \alpha_0^{\text{NTR}} G + \alpha G h(Z) + f(X;\alpha^{\text{NTR}}) + u^{\text{NTR}}, \tag{4}$$

where f is a function that is linear in parameters, and h is some function chosen by the researcher. We refer to the coefficient β as the *elasticity of taxable income* (ETI) estimand.

The TSLS model nests the triple differences model in equation (2) when $h(Z)$ is binary (equal to $\mathbb{1}[Z \neq 0]$) and f is a constant plus the indicator variable $\mathbb{1}[X \geq \bar{Y}]$. However, the TSLS model allows the researcher to choose a rich specification of f, thereby controlling flexibly for differential earnings- and tax rate growth, and to use variation in the intensity of treatment when the simulated instrument takes more than two values. The empirical ETI literature typically chooses $h(Z) = Z$ and specifies f to be a polynomial or spline function of X (see, e.g., Auten and Carroll (1999), Gruber and Saez (2002), and Kleven and Schultz (2014)).

2.3 Potential earnings model

To consider the causal interpretation of the ETI estimand (defined by equations (3) and (4)), it is necessary to introduce a potential earnings model that links the ETI estimand and the data to individual ETI parameters.

Given the tax systems T_0 and T_1, we let $Y_t(d)$ denote period t potential earnings under tax system T_d. Similarly, $\tau_t(d) \equiv T_d'(Y_t(d))$ denotes their potential marginal tax rate and $\text{NTR}_t(d) \equiv \log(1 - T_d'(Y(d)))$ their potential (log of net-of-) marginal tax rate. The potential outcomes map to observed outcomes through,

$$Y_t = \mathbb{1}[Z=0]Y_t(0) + \mathbb{1}[Z\neq 0]\left[(1-G)Y_t(0) + GY_t(1)\right], \tag{5}$$

$$\text{NTR}_t = \mathbb{1}[Z=0]\,\text{NTR}_t(0) + \mathbb{1}[Z\neq 0]\left[(1-G)\,\text{NTR}_t(0) + G\,\text{NTR}_t(1)\right], \tag{6}$$

for each $t \geq 0$, and $Y_t = Y_t(0), \text{NTR}_t = \text{NTR}_t(0)$ for $t < 0$.

Following Saez et al. (2012), we assume the potential earnings function is,

$$\log Y_t(\text{NTR}, d) = \zeta \times \text{NTR} + \nu_t(d), \tag{7}$$

where the parameters $\nu_t(0)$ and $\nu_t(1)$ can vary freely across individuals and ζ is the individual's ETI parameter.[5] Unless otherwise noted, we also allow ζ to vary freely across individuals. The specification in (7) allows the tax system to affect earnings both through marginal tax rates NTR and other channels ν.

It is useful to consider a set of assumptions commonly invoked to give instrumental variable estimands a causal interpretation, which imposes additional restrictions on the potential outcomes. Throughout the paper, we impose the following common trends assumption to recover the "reduced form" effects of the reform on earnings and the "first stage" effects of the reform on the marginal tax rates:

Assumption 1 (Common trends). *For each G, X, average earnings and marginal tax rates absent the tax reform changes according to*

$$\mathbb{E}\left[\Delta y(0) \mid G, X\right] = \lambda^y G + f_y(X), \tag{8}$$

$$\mathbb{E}\left[\Delta \,\text{NTR}(0) \mid G, X\right] = \lambda^{\text{NTR}} G + f_{\text{NTR}}(X), \tag{9}$$

where the functions f_y and f_{NTR} are unrestricted.

Assumption 1 states that average growth in earnings (marginal tax rates) conditional on initial income X and cohort G would have changed according to equation (8) (equation (9)) in the absence of the tax reform. The key restriction is that there is no interaction between G and X in the average growth in earnings (or marginal tax rates) in the absence of the reform. It means that the aggregate growth due to calendar-time effects (for example, due to business cycles) is allowed to vary freely over time, and that any idiosyncratic growth can vary freely across individuals depending on their initial income X.

Our second assumption is an exclusion restriction, which implies that individual earnings are affected by the tax reform only through its effects on marginal tax rates.

Assumption 2 (Exclusion restriction). $\nu_t(0) = \nu_t(1)$ *with probability one.*

As shown below, this restriction implies no income effects, an assumption we will relax in Section 4.

Lastly, we consider the assumption that the tax reform either weakly increases (or decreases) the marginal tax rates for all:

[5] The log specification implicitly assumes no extensive margin responses to the tax reform. In Section 3, we test this assumption empirically and find that the reform we consider did not affect extensive margin employment decisions.

Assumption 3 (Monotonicity)**.** $\mathbb{P}(\text{NTR}(1) \geq \text{NTR}(0)) = 1$ *or* $\mathbb{P}(\text{NTR}(1) \leq \text{NTR}(0)) = 1$.

This assumption is common in the program evaluation literature when allowing for treatment effect heterogeneity. It is typically necessary to ensure that standard IV estimands reflect positively weighted averages of individual treatment effects: see, e.g., Imbens and Angrist (1994).

2.4 Necessary and sufficient conditions for the ETI estimand to be causal

We now provide a characterization of the ETI estimand in terms of individual ETIs ζ under different choices of f and h, while maintaining the IV assumptions 1 - 3. A key goal is to understand when the ETI estimand β can(not) be given a causal interpretation as a positively weighted average of individual ETIs ζ:

Definition 1 (Causal ETI estimand)**.** *The ETI estimand β is* causal *if, for any distribution of individual ETIs ζ and initial income X, the maintained assumptions ensure that $\beta = \mathbb{E}[\omega \times \zeta]$ for some ω that satisfies $\mathbb{E}[\omega] = 1$ and $\mathbb{P}(\omega \geq 0) = 1$.*

The requirement that the weights sum to 1 ensures that a causal ETI estimand recovers ζ when ζ is constant across individuals, while the non-negative weights are necessary to ensure that a causal ETI estimand is contained in the support of ζ. For example, allowing for negative weights could mean that $\mathbb{E}[\omega \times \zeta]$ is negative even if ζ is always positive.

Proposition 1 provides necessary and sufficient conditions for the ETI estimand to be causal:

Proposition 1. *Suppose Assumptions 1 - 3 hold. Then, the ETI estimand β is causal if and only if $h(Z)$ is binary and f is unrestricted over the support of X.*

While we refer to Appendix A for a formal proof, it is useful to observe the three distinct reasons why an ETI estimand may fail to be causal.

First, if $h(Z)$ is non-binary, the ETI estimand will compare the earnings responses between individuals who experience (non-zero) changes in marginal tax rates of varying degrees. These comparisons could lead to negative weights if there is heterogeneity in the ETI parameters.[6] In contrast, if $h(Z)$ is binary, the ETI estimand only compares earnings changes between individuals treated by the reform ($Z \neq 0$) and untreated individuals ($Z = 0$), who do not experience any change in marginal tax rates. These comparisons will not produce negative weights, even if the elasticities are heterogeneous.

Second, if f is insufficiently flexible to capture how counterfactual earnings and marginal tax rate growth vary with X given G, then the excluded instrument $Gh(Z)$ in the first stage

[6] Callaway et al. (2025) formalizes the problem of variable treatment intensity in difference-in-differences, and shows that strong auxiliary assumptions are needed (e.g., constant effects) to give the estimate a causal interpretation.

(4) may be correlated with the error term u^y in the outcome equation (3). As a result, the ETI estimand cannot be given a causal interpretation, even when ζ is homogeneous.

If ζ is heterogeneous, the specification of f must also be flexible enough to ensure that it reproduces the conditional mean of the instrument, $\mathbb{E}[Gh(Z) \mid G = 1, X] = f(X)$. Specifications of f that are unrestricted are saturated in X and will always satisfy this condition.[7]

2.5 *The* causal interpretation of the ETI estimand

It is important to emphasize that the criterion of an estimand to be causal in Definition 1 is a weak one. Thus, being causal may be necessary for the ETI estimand to be an interesting quantity, but it is not sufficient. For example, the definition does not preclude that all the weight is assigned to a single person with a negative ζ, while the rest of the population receiving zero weight have positive ζ. The following corollary strengthens the result in Proposition 1 by showing *the* causal interpretation of the ETI estimand as a specific positively weighted average of individual ETI parameters ζ.

Corollary 1. *Suppose Assumptions 1 - 3 hold, that f is flexible and let $h(Z) = \mathbb{1}[Z \neq 0]$. Then, the ETI estimand β is causal and equals:*

$$\beta = \sum_{k=k_0}^{K} \sum_{j=1}^{J} \omega_{k,j} \times \underbrace{\mathbb{E}[\zeta \mid G = 1, X = x_k, \phi = \phi_j]}_{\text{group-specific average ETI}}. \tag{10}$$

where $\phi \equiv \text{NTR}(1) - \text{NTR}(0)$, $\{\phi_1, ..., \phi_J\}$ is the support of $\phi \mid \phi \neq 0$, $\{x_{k_0}, ..., x_K\}$ is the support of $X \mid X \geq \bar{Y}$, and the weights,

$$\omega_{k,j} = \frac{Var(G \mid X = x_k)\, \mathbb{P}(X = x_k, \phi = \phi_j \mid G = 1)\phi_j}{\sum_{l=k_0}^{K} Var(G \mid X = x_l) \sum_{m=1}^{J} \mathbb{P}(X = x_l, \phi = \phi_m \mid G = 1)\phi_m} \geq 0, \tag{11}$$

are positive, and sum to one.

The corollary shows that the ETI estimand recovers a specific positively weighted average of group averages of ζ. The groups are mutually exclusive and defined by initial income X and the reform's effect on their marginal tax change ϕ. Conditional on initial income $X = x$, groups are weighted according to their size and how large changes in the marginal tax rate ϕ they experience.

The ETI estimand aggregates the groups' average ETI parameters *across* initial income in proportion to how dispersed observations with $X = x$ are across cohorts G, as measured by $Var(G \mid X)$. These weights resemble how linear regression aggregates average treatment effects across covariates, and can be viewed as efficiency weights (See e.g., Angrist (1998)).

[7] Blandhol et al. (2022) show that the only specifications of the TSLS estimand that have a LATE interpretation are saturated specifications that control for covariates non-parametrically.

We now consider the special case when the ETI parameters ζ are assumed to be homogeneous across individuals. The following result shows that in this case, the ETI estimand recovers ζ provided f is unrestricted, thereby formalizing the identification argument implicit in the existing ETI literature:

Corollary 2. *Suppose Assumptions 1 and 2 are true, that ζ is constant across individuals, and that f is unrestricted over the support of X. Then, $\beta = \zeta$.*

The result is similar to Proposition 6 in Blandhol et al. (2022). Their linearity Assumption (LIN) is satisfied because of our Assumption 1, which ensures that the conditional mean of $\Delta y(0)$ is linear in X.

2.6 Quantifying how the elasticities vary across the income distribution

Corollary 1 showed that the ETI estimand β recovers a positively weighted average of individual ETI parameters ζ across groups defined by initial income X and marginal tax rate change ϕ. An important question for tax policy is how the ETI varies across the income distribution X. To analyze this question, we introduce the *local* ETI estimand $\beta(x)$, as defined by

$$\beta(x) \equiv \frac{\overbrace{\mathbb{E}[\Delta y \mid G=1, X=x]}^{\text{actual income growth}} - \overbrace{\left(\hat{\lambda}_0^y + \sum_{k=k_0}^{K} \hat{\lambda}_k^y \mathbb{1}[x_k = x]\right)}^{\text{counterfactual income growth}}}{\underbrace{\mathbb{E}[\Delta \, \text{NTR} \mid G=1, X=x]}_{\text{actual tax rate growth}} - \underbrace{\left(\hat{\lambda}_0^{\text{NTR}} + \sum_{k=k_0}^{K} \hat{\lambda}_k^{\text{NTR}} \mathbb{1}[x_k = x]\right)}_{\text{counterfactual tax rate growth}}}, \tag{12}$$

for each x in the support of $X \mid X \geq \bar{Y}$, where the coefficient vector $\hat{\lambda}^y$ ($\hat{\lambda}^{\text{NTR}}$) is obtained by regressing Δy (Δ NTR) on G and a set of dummies for each value of X for those with $GZ = 0$.

Intuitively, the estimand exploits that Assumption 1 implies the earnings and marginal tax rate growth among untreated ($GZ = 0$) individuals can be used to recover the counterfactual growth of the treated ($GZ \neq 0$) individuals. Subtracting the counterfactual growth from the treated individual's actual growth yields their earnings response and the changes in their marginal tax rate due to the tax reform. Our next result shows that the local ETI estimand $\beta(x)$ is causal under the same assumptions that were necessary to give the ETI estimand β a causal interpretation:

Proposition 2. *Suppose Assumptions 1 - 3 hold. Then, the local ETI estimand $\beta(x)$ is causal and equals:*

$$\beta(x) = \sum_{j=1}^{J} \overbrace{\frac{\mathbb{P}(\phi = \phi_j \mid G = 1, X = x)\phi_j}{\sum_{l=1}^{J} \mathbb{P}(\phi = \phi_l \mid G = 1, X = x)\phi_l}}^{\textit{weights reflecting group size and marginal tax rate effects}} \times \underbrace{\mathbb{E}[\zeta \mid G = 1, X = x, \phi = \phi_j]}_{\textit{group-specific average ETI}}, \tag{13}$$

for each $x \in \{x_{k_0}, ..., x_K\}$. The weights are positive and sum to one.

Proposition 2 shows that the local ETI estimand $\beta(x)$ recovers a positively weighted average across the same group-specific averages of ζ as in Corollary 1.[8] However, only groups with initial income $X = x$ receive positive weights. Among the groups with $X = x$, the local ETI estimand weights the groups according to their size $\mathbb{P}(\phi = \phi_k \mid G = 1, X = x)$ and how much their marginal tax rates are affected by the reform ϕ.

3 Estimating elasticities of taxable income

We now apply the identification results from Section 2 to analyze a reform of the Norwegian tax system that reduced the marginal tax rates on middle and high incomes.

3.1 The Norwegian tax system and the 2006 tax reform

Taxation of labor income. The Norwegian personal income tax combines a flat tax on general income with a progressive surtax on personal income.[9] General income consists of both labor and capital income. Labor income includes wages, salaries, and most employer-provided benefits. General income is taxed at a flat rate on net income, after allowable deductions. Deductions – such as the wage-earner deduction, the personal allowance, and selected pension and interest deductions – apply exclusively to the general-income tax base.

Personal income is defined as the sum of labor and pension income. It is subject to a 7.8 percent social security tax, which finances health and pension entitlements.[10] Personal income is also subject to a progressive surtax consisting of several income brackets with increasing marginal tax rates. The overall marginal tax rate on labor income thus combines the flat general income tax, the employee social security contribution, and the progressive surtax. Throughout the paper, we refer to the corresponding tax base as earnings.

[8] The expression is similar to equation (23) in Mogstad and Torgovitsky (2024) with the additional restriction that the potential earnings functions are linear. See their discussion of how it relates to the *average causal response* from Angrist and Imbens (1995).

[9] See Mogstad et al. (2025) for a detailed discussion of the Norwegian tax system and labor market.

[10] The rate varies depending on the source of the income, but is 7.8 percent for wage earners, which is the focus of our paper.

The 2006 tax reform. We study a tax reform that significantly reduced the surtax on middle- and high-income earners. The reform was partially introduced in 2005 and took full effect on January 1st, 2006. Figure 2 illustrates the marginal tax rate changes by plotting the marginal tax rates by personal income before (2004) and after (2006) the tax reform was fully implemented.[11] The graph shows that the reform reduced marginal tax rates on medium and especially high labor incomes: the surtax rates fell from 13.5 and 19.5 percent to 9 and 12 percent, respectively. This means that from 2004 to 2006, the marginal tax rate in the (second) highest bracket was reduced from 55.3 to 47.8 (49.3 to 44.8) percent. The sizable rate cuts meant that most workers in the upper half of the income distribution experienced large reductions in their marginal tax rates.

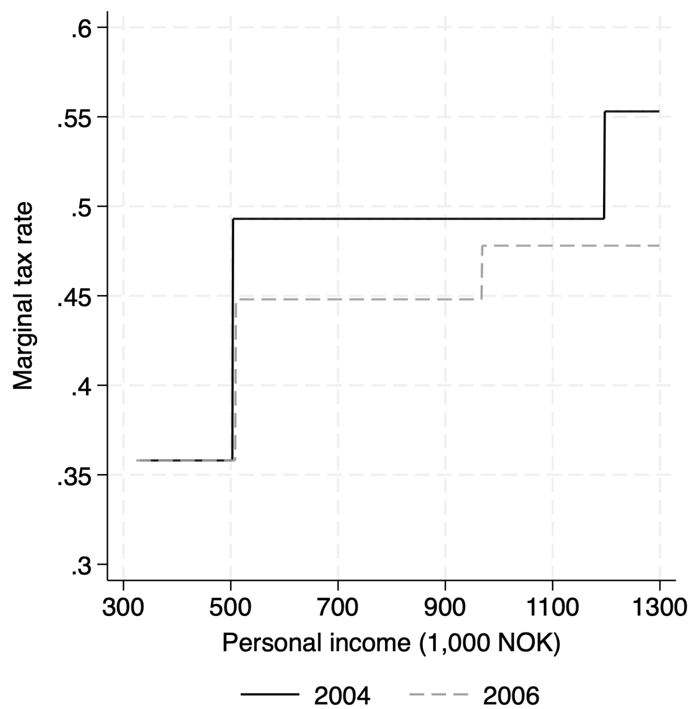


Figure 2: Marginal tax rates in 2004 and 2006

Notes: This figure shows the marginal tax rates that apply to different income tax brackets in 2004 and 2006. The tax bracket thresholds are measured in 2018 NOKs.

3.2 Data and sample

Our empirical analysis is based on several administrative data sources, which we link using unique identifiers for individuals and households. This results in a panel dataset covering the full Norwegian population in the period 1995–2018. The dataset includes detailed infor-

[11] Throughout the paper, we measure all monetary values in 2018 NOKs.

mation from income tax returns as well as individual characteristics such as age, sex, educational attainment, marital status, and number of children. Marginal tax rates are constructed using the same tax simulation model as Statistics Norway and the Norwegian Ministry of Finance (see Vattø (2020)).

Our baseline sample includes individuals aged 25 to 61 with wage earnings as their primary source of labor income. We exclude students and individuals receiving pensions or unemployment benefits.

Estimation sample. To explain how our estimation sample is constructed, it is useful to recall our discussion of the ETI estimands in Section 2. There, we showed that the ETI estimand compares how earnings and marginal tax rates evolve over periods when the tax system changed (2003 vs 2006) with how they evolve over periods when the tax system did not change (1999-2002). For each individual in the reform cohorts, we constructed one earnings difference by subtracting earnings in 2003 from earnings in 2006, and similarly for the marginal tax rates. Correspondingly, for each individual in the placebo cohorts, we constructed one earnings difference by subtracting earnings in 1999 from earnings in 2002.

It turns out that other year differences can also be used to analyze the reform. Because the reform was already (partly) implemented from January 1st, 2005, the differences between 2002 and 2005 and between 2004 and 2007 also capture how earnings and marginal tax rates evolve over periods where the tax system changed. Similarly, the differences between 2000 and 2003 and between 2001 and 2004 capture how earnings and marginal tax rates evolve when the tax system remained constant.

Our empirical analysis includes all six three-year differences. We refer to the differences over periods where the tax system changed (2002–2005, 2003–2006, and 2004–2007) as the reform differences. The differences over periods with no changes in the tax system (1999-2002, 2000–2003, and 2001–2004) are referred to as the placebo differences.

To obtain our estimation sample, we start from our baseline sample and construct the reform and placebo differences described above. We impose three additional restrictions. First, we exclude observations with initial income below 275,000 (15th percentile). Excluding individuals with low incomes is common practice in the literature since mean reversion tends to be most pronounced at the bottom of the income distribution (see Saez et al., 2012).

Second, we exclude observations with initial income above 950,000 (97th percentile). This restriction ensures that placebo differences are not affected by a top-income tax bracket introduced in 2000 for very high incomes (above 1,100,000).

Third, we exclude placebo differences with initial income between 375,000 and 450,000. This is done to minimize the extent to which placebo differences are affected by small year-to-year changes in tax bracket thresholds. Our estimates barely move if we include these individuals in the estimation sample.

Our resulting estimation sample contains about 4 million observations and 1.1 million individual wage earners. Summary statistics for our baseline and estimation samples are provided in Table 4 in Appendix B.

3.3 Estimates of average ETIs

To implement the ETI estimands, we pool the reform and control differences and estimate the model in equations (3) and (4) using two-stage least squares. Our specification includes separate dummies for each of the six three-year differences, and interacts the simulated instrument $h(Z)$ with a dummy G that equals 1 if the observation is a reform difference and zero if it is a placebo difference. Table 1 reports the resulting estimates under different choices of h and f.

The first column of Panel A reports estimates from our preferred specification, which puts $h(Z) = \mathbb{1}[Z \neq 0]$ and controls for initial income using a full set of percentile dummies in X. The first stage estimates reported in the first row are precise and show that the reform increases the marginal net-of-tax rate by around 5.3 percent. The F-statistic on the excluded instrument is above 70,000, indicating that the first stage is very strong. The reduced form, reported in the second row, is also precisely estimated and shows that the tax reform increases earnings by 1.2 percent.

The ETI estimate, reported in the third row, equals 0.233 (with a standard error of 0.012). The specification that produces this estimate controls for initial income non-parametrically and makes no comparisons of earnings responses between individuals who experience (non-zero) changes in marginal tax rates of varying degrees. Thus, it can be interpreted as a positively weighted average of ETI parameters under the assumptions in Proposition 1. If the ETI parameters are constant across individuals, the estimate implies that a ten percent increase in the marginal net-of-tax rate raises earnings by 2.3 percent.

The second and third columns of Panel A report estimates of specifications that fail to satisfy the conditions for a causal interpretation of the ETI estimand. The second column corresponds to the triple difference estimand in (2). It correctly specifies the instrument, but does not control flexibly for initial income, resulting in an estimate that understates the average ETI parameter.

The third column reports estimates from a commonly used ETI estimand that is causal only under the assumption of constant ETI parameters. It controls flexibly for initial income but specifies the instrument as $h(Z) = Z$. This means that it compares earnings responses between individuals who experience (non-zero) changes in marginal tax rates of varying degrees. The resulting estimate is equal to 0.161 and significantly understates the average ETI parameter, both statistically and economically. Thus, we can reject the assumption of homogeneous ETI parameters.

Table 1: Main results

Panel A. Specification of income controls f: dummies for X in bins.

Instrument		Binary: $h(Z) = \mathbb{1}[Z \neq 0]$		Continuous: $h(Z) = Z$
Income bins		Percentiles	Above/below $\bar{Y}$	Percentiles
First stage	Coef.	0.0526	0.0503	0.3045
	SE	0.0002	0.0002	0.0017
Reduced form	Coef.	0.0122	0.0100	0.0490
	SE	0.0006	0.0005	0.0039
ETI	Coef.	0.2325	0.1996	0.1609
	SE	0.0121	0.0109	0.0133

Panel B. Specification of income controls f: linear splines.

Instrument		Binary: $h(Z) = \mathbb{1}[Z \neq 0]$		Continuous: $h(Z) = Z$
Spline knots		Deciles	Above/below $\bar{Y}$	Deciles
First stage	Coef.	0.0527	0.0672	0.3009
	SE	0.0002	0.0002	0.0016
Reduced form	Coef.	0.0122	0.0111	0.0481
	SE	0.0006	0.0005	0.0038
ETI	Coef.	0.2306	0.1656	0.1599
	SE	0.0118	0.0076	0.0131

[a] The table contains two-stage least squares estimates of β from the model in (3) and (4) under different choices of h and f. All regressions include time dummies for each year and are weighted by initial income X.

Panel B of Table 1 reports estimates of specifications that control for initial income through splines instead of dummies for income bins. We report these estimates because spline specifications are frequently used in the ETI literature. We find that whether one flexibly controls for income through splines or income bins does not materially affect the results. We conclude that when the instrument is binary, and the income controls f are reasonably flexible, the ETI estimates are robust to the exact specification of f.

One possible concern with the validity of the estimates in Table 1 is that Assumption 1 (Common trends) may not hold. To address this concern, we re-estimate the model in equations (3) and (4) on the placebo differences, treating one of the placebo differences as if it were a reform difference. We perform this procedure for all three placebo differences and report the averages in Table 2.

Table 2: Placebo responses

Instrument		Binary: $h(Z) = \mathbb{1}[Z \neq 0]$		Continuous: $h(Z) = Z$
Income bins		Percentiles	Above/below $\bar{Y}$	Percentiles
Placebo reduced form	Coef.	0.00002	-0.00002	-0.06190
	SE	0.00001	0.00001	0.02200
Placebo ETI	Coef.	0.00042	-0.00047	-0.20340
	SE	0.00011	0.00011	0.07220

[a] This table presents placebo estimates of the reduced form from the model in equations (3) and (4). The placebo ETI is the reduced form divided by the actual first stage estimate from the corresponding specification in Table 1.

The resulting placebo reduced form, reported in the first column, is indistinguishable from zero. To obtain a placebo ETI estimate, one would need to divide the placebo reduced form by a first stage. However, because the placebo differences are not affected by any tax reform there is no first stage. We instead divide the placebo reduced form by the first stage estimate from the corresponding specification in Table 1. The resulting placebo ETI estimate is multiple orders of magnitude smaller than the actual ETI estimate.

Table 3: Employment responses

Instrument		Binary: $h(Z) = \mathbb{1}[Z \neq 0]$		Continuous: $h(Z) = Z$
Income bins		Percentiles	Above/below $\bar{Y}$	Percentiles
Employment response	Coef.	0.00060	0.00001	0.00270
	SE	0.00010	0.00010	0.00070

[a] This table presents employment estimates of the reduced form of the model in equations (3) and (4). The estimates are obtained by including unemployed individuals and replacing the outcome variable with a binary variable that is equal to one if the individual is employed.

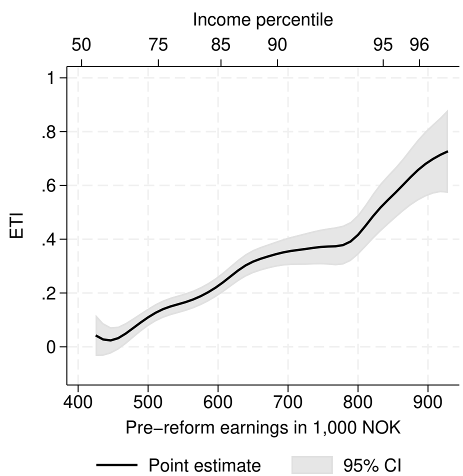
Another concern with the estimates in Table 1 is that the tax reform could induce extensive margin responses. To address this concern, we estimate the reduced form of the regression model in equations (3) and (4) on a sample that includes the unemployed, where we have replaced the outcome with a binary variable equal to one if the individual works and zero otherwise. Table 3 reports the resulting estimates. It shows that the tax reform had no meaningful impact on employment.

3.4 Estimates of ETIs across the income distribution

We now turn to assessing how the estimated ETI parameters vary across the distribution of initial income. We implement the estimator for $\beta(x)$ in two steps. The first step estimates the counterfactual income and tax rate growth in equation (12) using 100 percentile bins in X, while the second step estimates the numerator and denominator separately using a local regression.

Panel (a) of Figure 3 plots the resulting estimates. We find that the (nonlocal) ETI estimand considered above masks considerable heterogeneity. The average ETI parameter is less than 0.1 for incomes around 400,000 (median), increases steadily to around 0.35 for incomes close to 700,000 (90th percentile), and exceeds 0.5 for incomes around 850,000 (95th percentile). The variation in elasticities across the income distribution is significant, both statistically and economically, and means that the (weighted-) average elasticity recovered by the causal ETI estimand β is far from sufficient to assess how changes in marginal tax rates affect earnings and tax revenue.

(a) Local ETI.



(b) Local placebo ETI.

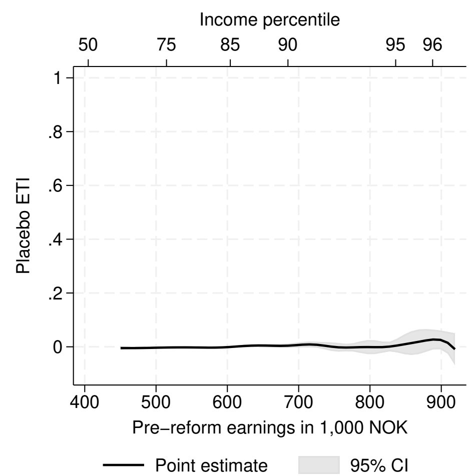


Figure 3: Local ETI estimates.

Notes: The figure plots actual and placebo estimates of the local ETI across the income distribution. Panel (a) plots the actual local ETI obtained by first estimating the counterfactual income and tax rate growth using 50 quantile bins of X, then estimating the numerator and denominator of equation (12) separately using a local regression. Panel (b) reports local placebo ETIs, obtained by estimating the numerator of equation (12) on the placebo cohorts only, sequentially treating each cohort as "treated," using the remaining two as controls, and averaging the resulting estimates. The local placebo ETI estimates are obtained by dividing these placebo reduced forms by the actual first stage from Panel (a). 95 percent confidence intervals are shown, with standard errors obtained by bootstrapping the entire estimation procedure using 500 replications.

Panel (b) of Figure 3 assesses the common trend assumption behind the local ETI estimand by estimating a placebo version of the local ETI estimand following the same steps as for the placebo estimates reported in Table 2. The placebo estimates remain close to zero throughout the income distribution. Figure 12 of Appendix B plots the local employment ETIs, which also remain indistinguishable from zero throughout the income distribution.

4 Using ETIs to learn about labor supply elasticities

We now use the ETI estimates from above to learn about labor supply elasticities.

4.1 A labor supply model

We consider a labor supply model where workers have convex preferences over K consumption goods C_k and L margins of labor supply Y_l. The L margins of labor supply produce pre-tax income Y according to $Y = F(Y_1, ..., Y_L)$, which is concave and strictly increasing in each of its arguments.

Multiple dimensions of labor supply allow for the possibility that individuals can affect their earnings through many different choices, including hours of work, effort on the job, and firm and occupation choices. Introducing multiple consumption goods is useful for computing total marginal tax rates, as it accommodates differential taxation, such as varying value-added tax rates, across goods.

If the income tax system were linear with marginal tax rate τ and transfer R, the individual's utility maximization problem would be,

$$\max_{C_1,...,C_K,Y_1,...,Y_L} U(C_1, ..., C_K, Y_1, ..., Y_L) \qquad \text{subject to } \sum_{k=1}^{K}(1+\tau_k)P_k C_k \leq I, \\ I = (1-\tau)Y + R + B, \text{ and } Y = F(Y_1, ..., Y_L), \tag{14}$$

where B is unearned income, I is consumption expenditure, and P_k and τ_k are the price of and tax on the consumption good k, respectively.

As noted by Feldstein (1999), the revenue- and efficiency effects of taxation typically depend on how tax policy affects earnings, but not on the specific margins through which individuals adjust their behavior. We therefore focus on the earnings choice, which – suppressing its dependence on prices and consumption tax rates – can be written as,

$$Y^u(\tau, R+B) \equiv F(Y_1^u(\tau, R+B), ..., Y_L^u(\tau, R+B)), \tag{15}$$

where $Y_l^u(\tau, R+B)$ is the optimally chosen l-th labor supply component. This earnings function $Y^u(\tau, R+B)$ allows for defining the standard labor supply elasticities,

$$\varepsilon^u \equiv \frac{\partial Y^u}{\partial 1-\tau}\frac{1-\tau}{Y^u}, \qquad \eta \equiv (1-\tau)\frac{\partial Y^u}{\partial(R+B)} \in [-1, 0], \qquad \varepsilon^c \equiv \varepsilon^u - \eta \geq 0, \tag{16}$$

where ε^u and ε^c denote the uncompensated and compensated earnings elasticity, respectively, η denotes the income effect, and the relationship between (un)compensated elasticities and the income effect is given by the Slutsky equation. The restriction that $\eta \in [-1, 0]$

follows from the Engel aggregation condition by assuming that consumption and leisure are normal goods.

In reality, the Norwegian tax system is piecewise linear. To accommodate this feature, we follow the argument of Hall (1973): convex preferences ensure that individuals behave as if they were facing the following linear budget constraint,

$$I = (1 - \tau(d))Y + R(d) + B \text{ for } d = 0, 1, \tag{17}$$

where

$$\tau(d) = T_d'(Y(d)), \qquad\qquad R(d) = T_d'(Y(d))Y(d) - T_d(Y(d)), \tag{18}$$

even if the actual tax system T_d is non-linear. This argument implies that the potential earnings function $Y(\tau(d), d)$ in equation (3) can be viewed as the solution to the worker's labor supply problem, subject to the linear budget constraint defined in equation (17):

$$Y(\tau(d), d) \ = \ Y^u(\tau(d), R(d) + B) \text{ for } d = 0, 1.$$

4.2 Recovering labor supply elasticities

The previous subsection forged a tight link between the potential earnings function and the model of labor supply. Having established this link, we now use the ETI estimates (and other specific features of the data) to draw inferences about labor supply elasticities.

We first consider the case with no income effects, a restriction typically imposed in the models used to interpret ETI estimands. Our next result shows that this model restriction ensures that the exclusion restriction in Assumption 2 is satisfied and, therefore, lets us interpret the individual ETI parameters ζ as compensated earnings elasticities ε^c:

Proposition 3. *If there are no income effects ($\eta = 0$ across all individuals), then Assumption 2 is satisfied and $\zeta = \varepsilon^c$ for each individual.*

An immediate implication of the result is that the ETI estimands β and $\beta(x)$ in Section 2 recover positively weighted averages of compensated earnings elasticities, provided the instrument is binary and the specification of the income controls f is unrestricted.

Allowing for income effects violates the exclusion restriction in Assumption 2, since the tax reform can then affect labor supply through changes in both the marginal and the average tax rates. The following proposition clarifies the identification problem that arises due to income effects. It expresses the compensated elasticities ε^c and the uncompensated elasticities ε^u in terms of the local ETI estimands $\beta(x)$ and a bias term that would be observable in data if the income effects were known or could be estimated:

Proposition 4. *Suppose that Assumption 1 holds, there is no bracket switching, and that ε^c and η are homogeneous among individuals with the same X. Then, for any $x \geq \bar{Y}$:*

$$\varepsilon^c(x) = \beta(x) + B(x)\eta(x), \tag{19}$$

$$\varepsilon^u(x) = \beta(x) + (B(x) - 1)\eta(x). \tag{20}$$

where $B(x)$ is estimable and equal to:

$$B(x) \equiv \frac{\mathbb{E}\left[\frac{T_1(Y_t)-T_0(Y_t)}{(1-T_1'(Y_t))Y_t} \mid G=1, X=x\right]}{\mathbb{E}[\Delta\,\mathrm{NTR} \mid G=1, X=x] - \left(\hat{\lambda}_0^{\mathrm{NTR}} + \sum_{k=k_0}^{K} \hat{\lambda}_k^{\mathrm{NTR}} \mathbb{1}[x_k = x]\right)}, \tag{21}$$

and $\hat{\lambda}^{\mathrm{NTR}}$ is defined as in Equation (12).

A key insight from Proposition 4 is that the local ETI estimand neither recovers the compensated elasticity nor the uncompensated elasticity if one allows for income effects, even if one assumes no bracket switching and that elasticities are homogeneous conditional on initial income X.[12] However, it also shows that the bias term is a multiplicatively separable function of the income effects $\eta(x)$ and an observable term $B(x)$. Therefore, if $\eta(x)$ were known or could be estimated, one could recover $\varepsilon^c(x)$ and $\varepsilon^u(x)$ from observed data. This observation motivates and guides the analysis in the remainder of the paper, where we consider different approaches to point identify or bound the compensated and uncompensated earnings elasticities.

4.3 No additional assumption bounds on labor supply elasticities

In order to use Proposition 4 to construct bounds on labor supply elasticities, it is useful to recall that the Engel aggregation condition implies that $\eta(x)$ is theoretically bounded between -1 and 0. Thus, evaluating equations (19) and (20) for $\eta(x) \in \{-1, 0\}$ produces bounds on compensated and uncompensated earnings elasticities without imposing assumptions other than those stated in Proposition 4.

The light grey areas of Figure 4 plot these bounds using the estimates of $\beta(x)$ reported in Figure 3 and estimates of $B(x)$ from the cross-sectional earnings distribution. Panel (a) provides two insights about compensated elasticities. First, the compensated elasticities for individuals with incomes equal to 600,000 (85th percentile), 700,000 (90th percentile), and 900,000 (96th percentile) are at least $0.2, 0.35$, and 0.7, respectively. Second, the bounds rule out larger than modest compensated elasticities for middle-income individuals: the upper

[12]The result in Proposition 4 does not invoke Assumption 3. It is no longer needed since it is assumed that elasticities are homogeneous among individuals with the same initial income X. For tractability, Proposition 4 also imposes no bracket switching, which will necessarily hold if the reform under consideration is only changing the top income tax.

bound for individuals with income around 400,000 (median) equals 0.5. Taken together, these insights imply that compensated elasticities increase by at least 50 percent as income increases from the median to the 96th percentile.

Panel (b) plots the corresponding bounds for the uncompensated elasticities. A key result is that the bounds are highly informative for high-income individuals, implying a relatively large uncompensated elasticity between 0.4 and 0.7 for individuals with incomes at the 95th percentile. A second result is that we can rule out that income effects dominate substitution effects for all individuals with income above 700,000 (90th percentile). Finally, the bounds rule out moderately positive uncompensated elasticities: individuals with incomes below 500,000 (75th percentile) have an uncompensated elasticity below 0.1.

(a) Compensated elasticity. (b) Uncompensated elasticity.

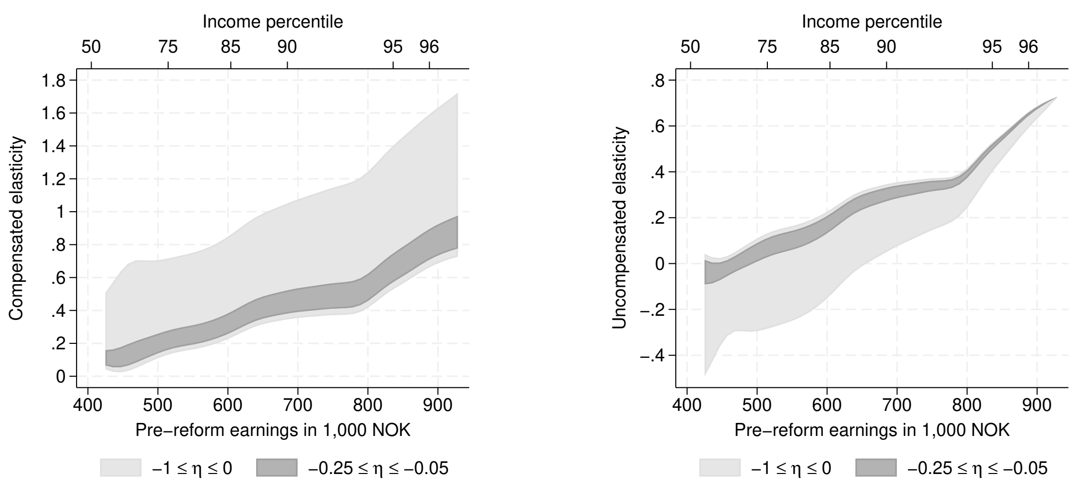


Figure 4: Average (un)compensated earnings elasticities across initial income X.

Notes: The figure plots bounds on the compensated and uncompensated labor supply elasticities over the income distribution. The light grey area in Panel (a) plots the bounds on $\varepsilon^c(x)$ using that $\eta \in [-1, 0]$, while the darker grey area imposes that $\eta \in [-0.25, 0.05]$. The light grey area in Panel (b) plots the bounds on $\varepsilon^u(x)$ using that $\eta \in [-1, 0]$, while the darker grey area imposes that $\eta \in [-0.25, 0.05]$.

These bounds can be tightened considerably by ruling out implausibly small and large income effects. The dark grey areas in Figure 4 represent the bounds implied by assuming that $\eta \in [-0.25, -0.05]$. They show that this additional assumption tightens both the upper bound on the compensated elasticities and the lower bound on the uncompensated elasticities considerably. The implied bounds suggest an (un)compensated elasticity between 0.1 and 0.2 (between -0.1 and 0) for incomes around 450,000. The (un)compensated elasticity is bounded between 0.4 and 0.5 (0.25 and 0.35) for incomes around 700,000.

Although the bounds presented above are highly informative, the (un)compensated elasticities remain partially identified. We next consider two ways to move from partial to point identification. One possibility, which we consider in subsection 4.4, is to assume constant income and substitution effects across the income distribution. Another possibility, which we consider in Section 5, is to combine the ETI estimands with external estimates of income effects.

4.4 Point identification with constant income and substitution effects

As shown by Gruber and Saez (2002), it is possible to use variation from tax reforms to jointly estimate income and substitution effects under the assumption that the compensated elasticity ε^c and the income effect η are constant across all individuals. Assuming homogeneity, Proposition 4 implies that the local ETI estimands $\beta(x)$ relate to the constant ε^c and η through,

$$\beta(x) = \varepsilon^c + \eta B(x). \tag{22}$$

A natural way to estimate (ε^c, η) is the following least squares estimator

$$(\hat{\varepsilon}^c, \hat{\eta}) = \underset{\varepsilon^c, \eta}{\arg\min} \sum_{k=k_0}^{K} p_k \left(\hat{\beta}(x_k) - \varepsilon^c - \eta \hat{B}(x_k) \right)^2 \tag{23}$$

where $p_k \equiv \mathbb{P}(X = x_k \mid G = 1)$ are weights corresponding to each point in the support of $X \mid X \geq \bar{Y}$. An estimator for the uncompensated elasticity is then $\hat{\varepsilon}^u = \hat{\varepsilon}^c + \hat{\eta}$.

Since equation (22) must hold for any x, ε^c and η are (over)identified if $\beta(x)$ and $B(x)$ are observed for (more than) two values of x. Even with only two values of x, the homogeneity assumption can be tested by examining if the labor supply parameters satisfy the theoretical restrictions $\varepsilon^c > 0$ and $\eta \in [-1, 0]$. In the overidentified case, the sharp test would not only use these theoretical restrictions, but also that the sum of squared residuals in (23) is zero.

Using the same estimates of $\beta(x)$ and $B(x)$ as in Figure 4, we solve the problem in (23) and obtain $\hat{\varepsilon}^c = -0.64$ and $\hat{\eta} = 1.32$. Since the estimated compensated elasticity is negative and the income effect is positive, the homogeneity assumption is clearly at odds with the data. In fact, we can statistically reject the assumption at any reasonable level of significance.

To understand why the assumption is rejected, it is useful to recall that the tax reform we consider decreased the top income tax rate and, as a result, the reduction in average tax rates is increasing in income. The assumption of constant ε^c and η therefore implies that earnings responses should decrease across the income distribution, in sharp contrast with Figure 3. Thus, we conclude that the assumption of constant income and substitution effects poorly approximates individual labor supply behavior, at least in our context.

5 Combining ETIs with external estimates of income effects to learn about labor supply elasticities

Proposition 4 shows that the compensated $\varepsilon^c(x)$ and uncompensated elasticities $\varepsilon^u(x)$ can be point identified from the local ETI estimands $\beta(x)$ if income effects $\eta(x)$ are known or can be estimated externally for each income level x. Motivated by this result, we now use Norwegian data on lottery winnings to measure income effects. We begin this analysis by estimating earnings and employment responses to lottery-induced changes in unearned income. Next, we show how these responses allow us to infer the income effects on the intensive margin that we need to point identify the (un)compensated elasticities.

5.1 Data and sample

Our empirical analysis combines multiple administrative data sources linked through unique personal and household identifiers. We supplement the tax records on wealth with measures of market values of real estate, using data on transactions in real estate, information on the characteristics of each property, and detailed housing price indices. The resulting panel covers the entire Norwegian population from 1995 to 2018 and includes demographic characteristics, detailed tax records on income and wealth, and information on lottery prizes and asset values.

We restrict our sample to winners aged 25–61 in the year before the win and exclude students and individuals receiving pensions or unemployment benefits. Since individuals are only required to report winnings of NOK 100,000 or more, we restrict our sample to those who won at least this amount.[13] The final sample includes more than 14,000 unique winners across 23 years. The median price is 343,000 NOK.

Table 5 in Appendix B compares working-age lottery winners with the general population. It shows that winners are somewhat older, more often male, and have slightly higher earnings. However, the two groups have similar employment rates, years of schooling, and are equally likely to be married. By and large, we find that lottery winners are broadly similar to the working-age population.

[13] This threshold has been in place since 2007. Before 2007, the threshold was NOK 10,000.

5.2 Measuring unearned income

To understand how we measure period-by-period unearned income, it is useful to start with the household's intertemporal budget constraint,

$$C_t = Y_t - T(Y_t) + \underbrace{\Big((1+r)A_t - A_{t+1}\Big)}_{\text{unearned income} \equiv B_t}, \tag{24}$$

where C_t is period t consumption expenditure, A_{t+1} is assets held at the end of period t, r is the net-of-tax return rate on the asset, and B_t is the total amount of unearned income used by the household in period t, or unearned income for short.[14] Lottery winnings provide an exogenous increase in unearned income, and, as we show in Appendix E, the earnings responses to this variation are directly linked to the individuals' income effects.

Following Eika et al. (2020), we use our detailed data to construct household-level measures of consumption and savings. These measures allow us to compute unearned income directly using equation (24), which we then convert to a per-adult measure for consistent comparison between single and married households. This means that our rich data allow us to *observe* how winners allocate their wealth over time, eliminating the need to rely on the annuitization or capitalization approaches used in Imbens et al. (2001) and Golosov et al. (2024).

5.3 Research design and estimation

Before describing our research design, it is useful to introduce some terminology. We call all individuals who won a lottery in a given calendar year g a cohort, and denote an individual's cohort by G. The event time t for cohort g corresponds to calendar year $g + t$, where t can be positive or negative depending on whether we look at the outcomes before or after winning a lottery. We use the year before the lottery win as the reference year in the event study and refer to this year as the pre-win year. Subscripts now denote calendar year.

We explain our research design by showing how we recover the earnings effect of winning a lottery prize. We employ a difference-in-differences design that compares the evolution of earnings over time for individuals who have already won the lottery with those who have yet to win. For each event time t, we compare the earnings changes between years $g-1$ and $g+t$ for cohort g to changes over the same two years for cohorts that will receive the prize at a later date, so that the comparison group remains untreated at both points in time.

[14] This definition of unearned income is consistent both with intertemporal two-stage budgeting in the absence of liquidity constraints and with the presence of liquidity constraints (Arellano and Meghir, 1992; Blundell and Macurdy, 1999; Blundell and Walker, 1986 and MaCurdy, 1983).

Formally, we define the difference-in-differences estimand by:

$$\mathrm{RF}_{g,t}(x) \equiv \underbrace{\mathbb{E}\left[Y_{g+t} - Y_{g-1} \middle| G = g, Y_{g-1} = x\right]}_{\text{earnings change for year } g \text{ winners}} - \overbrace{\mathbb{E}\left[Y_{g+t} - Y_{g-1} \middle| G > g + t, Y_{g-1} = x\right]}^{\text{earnings change for later-than } g+t\text{-winners}}. \tag{25}$$

This estimand captures how earnings change for lottery winners relative to individuals with the same pre-win earnings Y_{g-1} who have not yet won. We show in Appendix C that $\mathrm{RF}_{g,t}(x)$ where $t \geq 0$ recovers the earnings effect of winning the lottery under a standard parallel trends assumption.

In estimation, we implement the conditioning on Y_{g-1} using kernel weights and estimate $\mathrm{RF}_{g,t}(x)$ separately for each cohort g and event time t. The estimands $\mathrm{RF}_{g,t}(x)$ are then aggregated across cohorts using cohort-size weights. When aggregating across event time, we weight each event time equally. For each x, we estimate all cohort-by-event-time parameters in a single, fully interacted specification, which allows us to compute standard errors for aggregated effects using the delta method. All specifications include flexible controls for age to account for any systematic age differences between earlier and later winners.

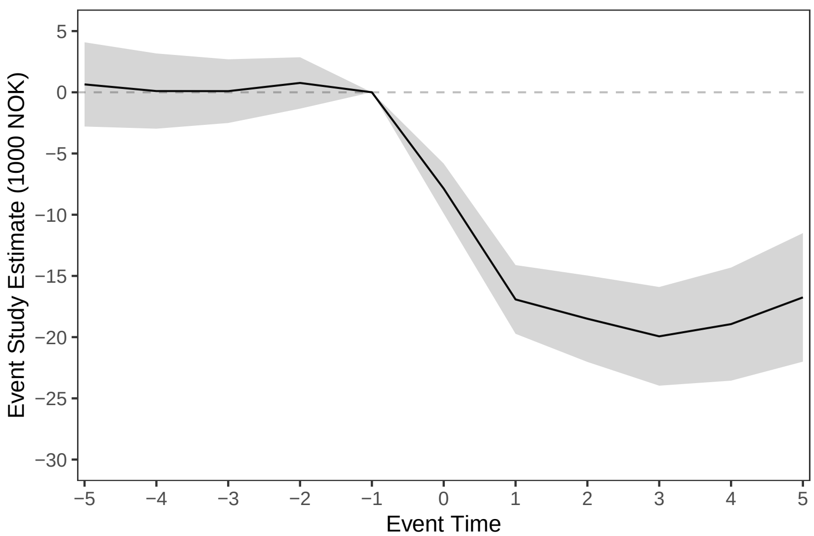
5.4 The effect of winning the lottery on earnings and employment

Panel (a) of Figure 5 presents unconditional estimates of the effect of winning on winners' earnings for different event times. The estimates show no evidence of systematically different pre-trends between current and future winners. They reveal that earnings decline sharply in the first two years after winning and then stabilize at a lower level: After 3 years, average earnings are reduced by around NOK 20,000.

The average effects in Panel (a) could mask considerable heterogeneity. Panel (b) shows how pre- and post-win estimates vary across the distribution of pre-win earnings and shows that pre-trends are similar between treatment and controls throughout the distribution. Following the win, we observe a substantial earnings reduction of around 15,000-25,000 that increases in magnitude with pre-win income.

Figure 13 in Appendix B plots the corresponding figures for the employment response. Panel (a) shows no evidence of systematic differences in employment time trends between current and later winners. After winning, the employment of winners declines over time and is reduced by 1.5 percentage points after 5 years. Panel (b) plots the pre- and post-event employment response estimates across the distribution of pre-win earnings. The employment responses are more pronounced at lower levels of pre-win income. While winning reduces employment by 1.25 percentage points among winners with pre-win earnings in the 10th percentile, the corresponding response among winners with pre-win earnings in the 90th percentile is 0.5 percentage points.

(a) Unconditional effect on earnings



(b) Conditional effect on earnings

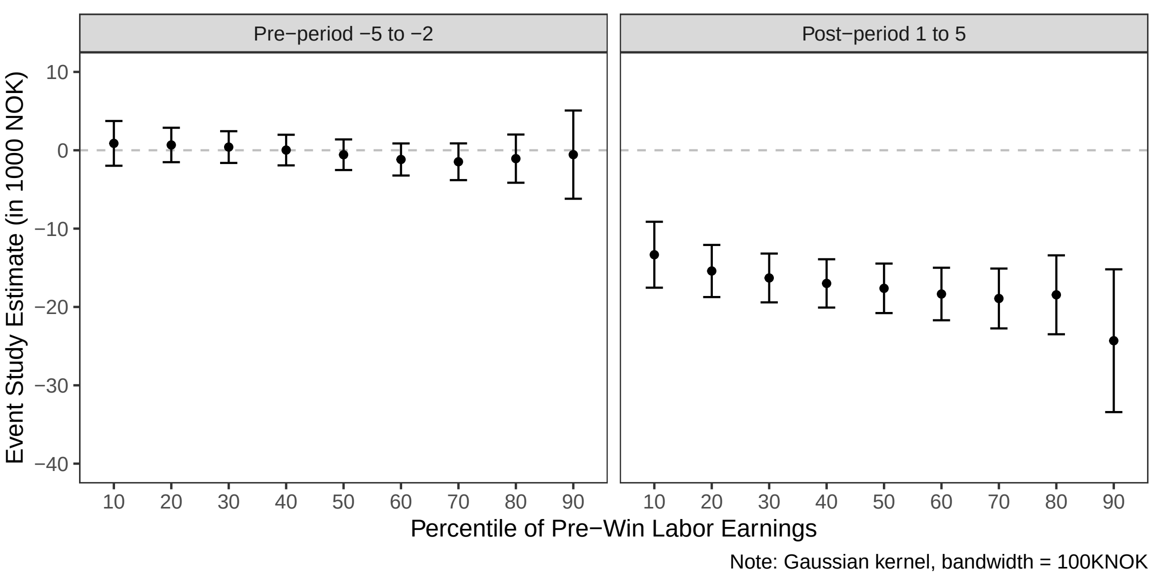


Figure 5: Earnings effects of winning

Notes: This figure presents estimates of the effect of winning the lottery on winners' earnings. Panel (a) presents unconditional estimates of the effect of winning the lottery on winners' earnings for each event time t. The estimates correspond to the sample analogue of equation (25), evaluated without conditioning on X, controlling for age, and averaged across cohorts using cohort-size weights. The estimates in Panel (b) correspond to the sample analogs of equation (25), evaluated at different values of X. We use a Gaussian kernel with a bandwidth of 100,000 NOK, where the evaluation points $\{x\}$ correspond to the cohort-specific deciles of pre-win earnings. For each decile, we compute cohort-size weighted averages across cohorts and report equally weighted averages across event times $t = -5$ to -2 and $t = +1$ to $+5$. 95 percent confidence intervals are shown, with standard errors clustered at the winner level. Throughout, we use $g - 1$ as the omitted event time.

5.5 The earnings response to unearned income

The size of the effects that we estimated in the previous section can be hard to gauge, as the observed responses to windfall gains could vary across individuals depending on a number of factors, such as the age at which the individual wins and her savings behavior. Motivated by this, we now turn to estimating how earnings respond to plausibly exogenous variation in unearned income generated by lottery winnings.

Using the same difference-in-differences design as above, we first recover the effect of winning the lottery on unearned income by considering the following estimand,

$$\text{FS}_{g,t}(x) \equiv \underbrace{\mathbb{E}\left[B_{g+t} - B_{g-1} \middle| G = g, Y_{g-1} = x\right]}_{\text{unearned income change for year } g \text{ winners}} - \overbrace{\mathbb{E}\left[B_{g+t} - B_{g-1} \middle| G > g+t, Y_{g-1} = x\right]}^{\text{unearned income change for later-than } g+t\text{-winners}}, \quad (26)$$

for each g, t and x. We estimate and aggregate $\text{FS}_{g,t}(x)$ across cohorts and event times as we did above.

Panel (a) of Figure 6 presents the resulting estimates. It shows no evidence of different pre-trends between winners and later winners in any part of the distribution. Winning the lottery increases unearned income for all income levels. The effect equals around 50,000 for the lowest decile and gradually increases to around 80,000 in the highest decile.

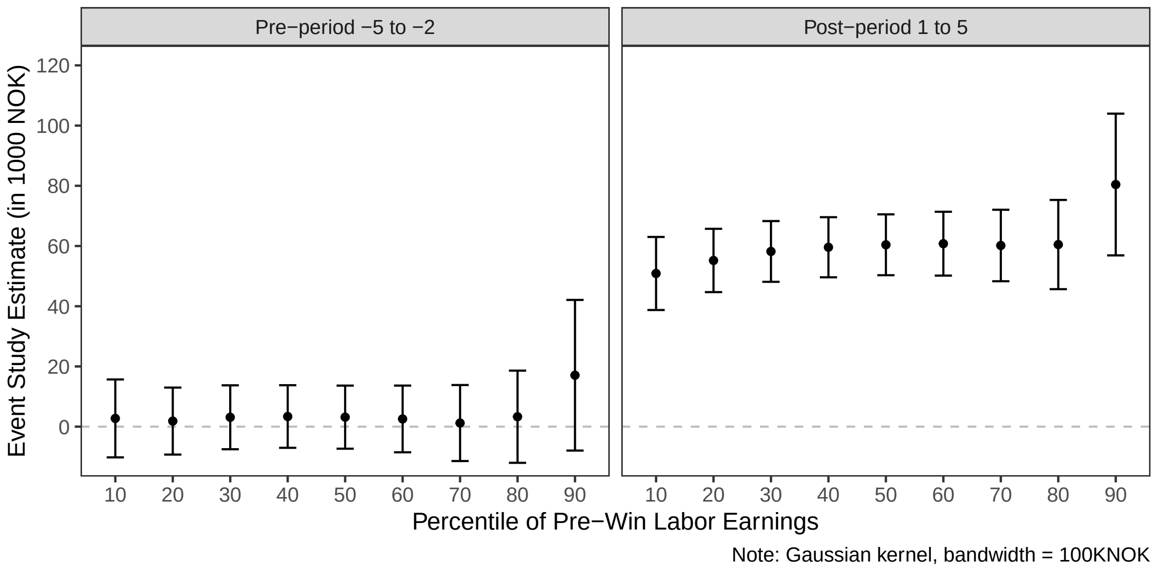
To obtain the earnings response to unearned income, we take the ratio between the aggregated reduced-form and first-stage estimates. Panel (b) of Figure 6 reports these IV estimates across the distribution of pre-win earnings. It shows that the estimates are remarkably stable across the earnings distribution. On average, an additional NOK of unearned income reduces earnings by 0.3. This number aligns closely with the estimates for the bottom quartile from Golosov et al. (2024), but is lower than their reported average of 0.5, mainly because higher-income US households are more responsive.

5.6 Recovering the intensive-margin income effect

The estimates in Figure 6 reflect a combination of intensive and extensive-margin labor supply responses, while the parameter we need to point identify (un)compensated elasticities ($\eta(x)$) is the intensive-margin labor supply response to changes in unearned income. To recover intensive-margin income effects, we first decompose the total earnings response to unearned income into its intensive- and extensive-margin components.[15] The decomposition relies on an additional common-trend assumption: conditional on pre-win earnings, winners who stay employed would have experienced the same earnings trend as later winners who also remain employed, had the former not won.

[15]See Appendix C for the formal derivation of the decomposition.

(a) Effect on unearned income



(b) Earnings effect of unearned income

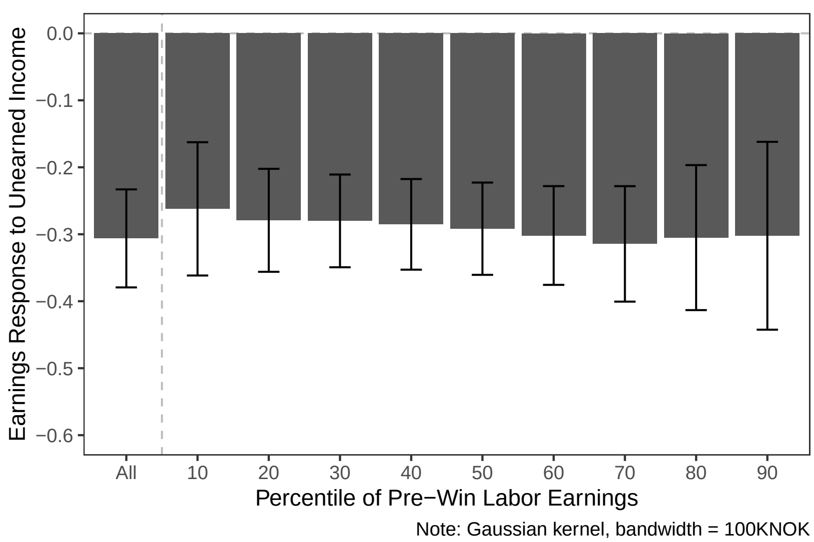


Figure 6: Unearned income effects of winning and IV estimates

Notes: This figure presents estimates of the effect of winning the lottery on winners' unearned income and the earnings response per unearned income. The estimates in Panel (a) correspond to the sample analogs of equation (26), evaluated at different values of X. We use a a Gaussian kernel with a bandwidth of 100,000 NOK, where the evaluation points $\{x\}$ correspond to the cohort-specific deciles of pre-win earnings. For each decile, we compute cohort-size weighted averages across cohorts and report equally weighted averages across event times $t = -5$ to -2 and $t = +1$ to $+5$. The estimates in Panel (b) correspond to the ratio between the aggregated values of $\text{RF}_{t,g}(x)$ and $\text{FS}_{t,g}(x)$ evaluated at different values of x. 95 percent confidence intervals are shown, with standard errors clustered at the winner level. Throughout, we use $g - 1$ as the omitted event time.

(a) Unearned income response decomposition. (b) Income effect.

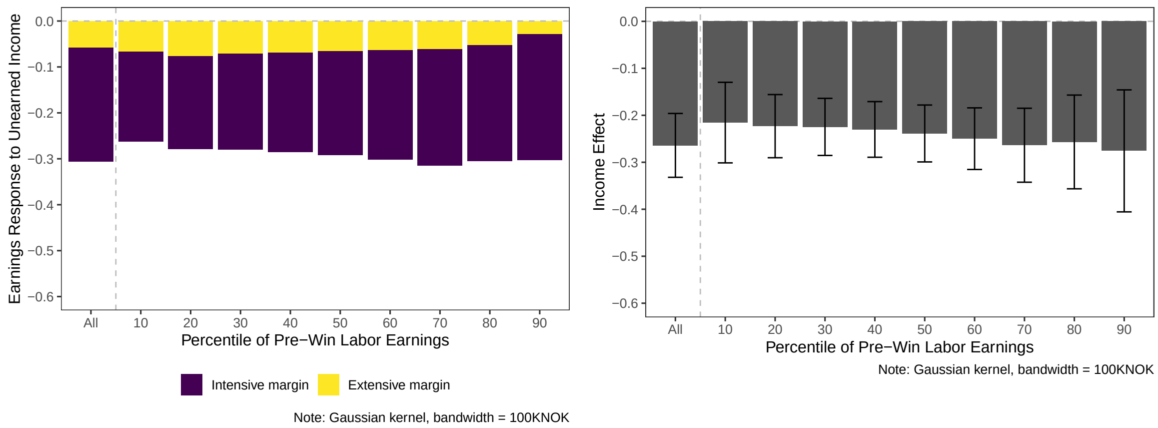


Figure 7: Total and intensive-margin income effects.

Notes: Panel (a) decomposes the earnings response per extra NOK of unearned income into its intensive- and extensive- margin components according to Equation (67). Panel (b) plots the intensive-margin income effects conditional on pre-win labor earnings. 95 percent confidence intervals are shown, with standard errors clustered at the winner level. Throughout, we use $g-1$ as the omitted event time.

Panel (a) of Figure 7 plots the results from the decomposition. It shows that the intensive-margin component is by far the most important one, accounting for 80-85 percent of the total response. The intensive-margin share increases with income and accounts for more than 95 percent of the total response for individuals with pre-win earnings at the 90th percentile.

To obtain the intensive-margin income response, it is useful to note that the intensive-margin contribution equals the share of individuals who respond on the margin multiplied by their response. To obtain the intensive-margin response, we therefore divide the intensive-margin contribution by the share that responds on the intensive margin, i.e., the share that continues to work after winning.

Panel (b) plots the resulting intensive-margin responses. It shows that the intensive-margin earnings response to an additional NOK of unearned income is, on average, -0.25. This estimate is stable across the distribution of pre-win earnings.

Finally, the income effects $\eta(x)$ that we need to point identify the (un)compensated elasticities can now be inferred by multiplying the intensive-margin response reported in Panel (b) by the net-of-tax rate at the corresponding earnings level.

5.7 Point identification with external estimates of income effects

Figure 8 presents point estimates of compensated and uncompensated earnings elasticities obtained using our estimated income effects. It shows that both elasticities are monotonically increasing in income. Panel (a) shows a compensated elasticity of 0.1 for incomes around 400,000 NOKs (median) that increases steadily to around 0.3 for incomes close to 700,000

(90th percentile), and equals 0.7 for incomes around 850,000 (95th percentile). On average, the compensated elasticity equals 0.34.

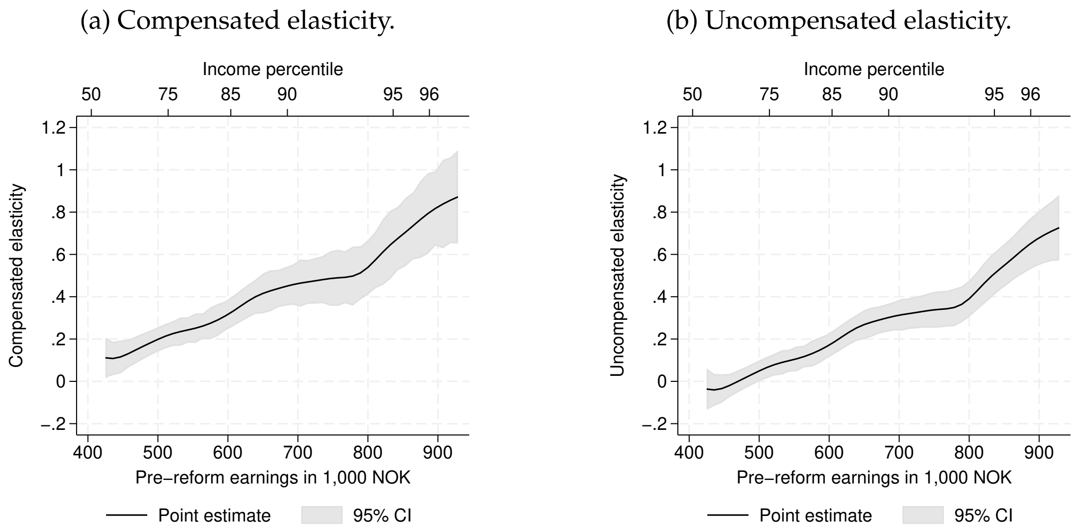


Figure 8: Average (un)compensated earnings elasticities across initial income X.

Notes: The figure plots point estimates of the compensated (Panel (a)) and uncompensated (Panel (b)) labor supply elasticities over the income distribution. 95 percent confidence intervals are shown, with standard errors obtained by assuming that the lottery-based income effects $\eta(x)$ and local ETIs $\beta(x)$ are uncorrelated.

Panel (b) shows that the uncompensated elasticity is close to zero for individuals with income between 400,000 and 500,000, both statistically and economically. For incomes above the 75th percentile, the uncompensated elasticity is positive and significantly different from zero. It equals 0.3 for incomes around 700,000 (90th percentile) and increases steadily to around 0.6 for incomes around 850,000 NOKs (95th percentile). On average, the uncompensated elasticity equals 0.19. In conclusion, substitution effects tend to be larger than income effects, especially for higher incomes.

6 Revenue-maximizing tax rates and excess burden

In this section, we use our estimates of income and substitution effects to quantify the efficiency cost of increasing the marginal tax rates on middle- and high-income individuals. We also calculate the implied revenue-maximizing top-income tax rate. To do so, we begin by tailoring the standard expressions for marginal deadweight loss (Auerbach and Hines, 2002 and Harberger, 1964) and the revenue-maximizing top income tax rate (Diamond, 1998 and Saez, 2001) to the Norwegian institutional context by accounting for payroll and value-added taxes. Next, we evaluate these expressions using our estimates of income and substitution effects from Section 5.

6.1 Measuring the effective tax rate

To derive expressions for the excess burden and revenue-maximizing top-income tax rates, it is necessary to account for the value-added and payroll taxes that are part of the Norwegian tax system. To this end, we suppose the tax revenue collected from an individual with a consumption bundle $(C_1, ..., C_K)$ and earnings Y equals,

$$TR(C_1, ..., C_K, Y) \equiv \underbrace{\sum_{k=1}^{K} \tau_k P_k C_k}_{\text{consumption tax}} + \overbrace{T(Y)}^{\text{income tax}} + \underbrace{\tau_w Y}_{\text{payroll tax}},$$

where τ_w is the proportional payroll tax rate and τ_k is the tax rate on consumption good k.

The tax rates τ_k vary across consumption goods. This means that to calculate the revenue effect of changes in the income tax, one would need to know how it affects the demand for each of the K consumption goods. To reduce the dimensionality of the problem, we specialize our model from Section 4 by assuming that individual preferences are separable between the consumption and labor supply components:

$$U(C_1, ..., C_K, Y_1, ..., Y_L) = \underbrace{U_c(C_1, ..., C_K)}_{\text{utility from consumption}} + \overbrace{U_y(Y_1, ..., Y_L)}^{\text{disutility from labor}}.$$

By standard two-stage budgeting arguments, this means the uncompensated demand for the k-th consumption good depends only on the income tax through its effect on disposable income:

$$C_k = C_k^u\Big(\tau_1, ..., \tau_K, \underbrace{I^u(\tau_1, ..., \tau_K, \tau, R + B)}_{\equiv(1-\tau)Y^u(\tau_1,...,\tau_K,\tau,R+B)+R+B}\Big). \tag{27}$$

Using these demand functions, we can express the effective consumption tax rate as

$$\tilde{\tau} \equiv \frac{\sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^u}{\partial I}}{\sum_{k=1}^{K} P_k \frac{\partial C_k^u}{\partial I}}. \tag{28}$$

As we show ahead, this tax rate is a sufficient statistic for calculating how changes in the income tax system affect the revenue raised from consumption taxes.

By assuming that each consumption good is normal, we obtain $\tilde{\tau}$ is bounded between the lowest and largest consumption tax rate. In our setting, this means that the effective consumption tax rate is bounded between 0 and 0.25. We allow for any effective tax consumption tax rate in this range but impose that it does not vary across individuals.[16] Using average expenditure shares to proxy for the marginal expenditure shares implies that $\tilde{\tau} = 0.21$.

6.2 Excess burden of taxation

We now quantify the efficiency cost of increasing the marginal tax rate on middle and high-income individuals by introducing the excess burden.

Deriving the excess burden. To derive a measure of the excess burden of taxation, it is useful to consider the individual's expenditure minimization problem:

$$\begin{aligned} E(\tau_1, ..., \tau_K, \tau, \bar{V}) \equiv \min_{C_1,...,C_K,Y_1,...,Y_L} & \sum_{k=1}(1+\tau_k)P_k C_k - (1-\tau)F(Y_1, ..., Y_L) \\ & \text{subject to } U(C_1, ..., C_K, Y_1, ..., Y_L) \geq \bar{V}, \end{aligned} \tag{29}$$

where $\bar{V}$ is individual utility after the tax reform. The resulting compensated demand functions can be written as $C_k^c(\tau_1, ..., \tau_K, \tau, \bar{V})$ and the earnings function as $Y^c(\tau_1, ..., \tau_K, \tau, \bar{V}) \equiv F(Y_1^c(\tau_1, ..., \tau_K, \tau, \bar{V}), ..., Y_L^c(\tau_1, ..., \tau_K, \tau, \bar{V}))$.

These functions allow us to define the deadweight loss of taxation,

$$DWL(\tau_1, ..., \tau_K, \tau, \bar{V}) \equiv E(\tau_1, ..., \tau_K, \tau, \bar{V}) - TR^c(\tau_1, ..., \tau_K, \tau, \bar{V}),$$

where the tax revenue TR^c is given by,

$$TR^c(\tau_1, ..., \tau_K, \tau, \bar{V}) \equiv \sum_{k=1}^{K} \tau_k P_k C_k^c(\tau_1, ..., \tau_K, \tau, \bar{V}) + (\tau + \tau_w) Y^c(\tau_1, ..., \tau_K, \tau, \bar{V}).$$

In words, the deadweight loss is the additional expenditure, beyond the revenue raised, that is required to keep an individual at utility level $\bar{V}$ under the tax system $(\tau_1, ..., \tau_K, \tau)$.

The marginal deadweight loss is obtained by taking the derivative of DWL with respect to τ. The excess burden EB normalizes the marginal deadweight loss by the marginal revenue raised, i.e., the derivative of TR^c with respect to τ:

$$EB \equiv \frac{\mathbb{E}\left[\frac{\partial DWL}{\partial \tau} \mid Y \geq \bar{Y}\right]}{\mathbb{E}\left[\frac{\partial TR^c}{\partial \tau} \mid Y \geq \bar{Y}\right]}. \tag{30}$$

In words, the excess burden EB measures the economic cost of increasing the marginal tax rate slightly per additional NOK of revenue raised.

[16] This is true if, for example, the consumption component of the utility function U_c is homothetic and homogeneous across individuals.

We show in Appendix D.1 that the excess burden can be expressed in terms of compensated elasticities, tax rates, and earnings as:

$$EB = \frac{\mathbb{E}\left[\left(\frac{\tilde{\tau}}{1+\tilde{\tau}} + \frac{\tau+\tau_w}{1-\tau}\right)\varepsilon^c Y \mid Y \geq \bar{Y}\right]}{\mathbb{E}\left[\left(1-\left(\frac{\tilde{\tau}}{1+\tilde{\tau}} + \frac{\tau+\tau_w}{1-\tau}\right)\varepsilon^c\right) Y \mid Y \geq \bar{Y}\right]}. \tag{31}$$

Quantifying the excess burden of taxation. We now calculate the excess burden in equation (31) under two different sets of estimates of elasticities. The first set is obtained from our data and estimates of ε^c from Section 5, reported in Panel (a) of Figure 8. We refer to these estimates as our preferred specification. The second is the compensated elasticities one would obtain from using the common ETI estimand, ignoring income effects, and assuming constant labor supply elasticities. Under this conventional specification, $\varepsilon^c = 0.16$ and $\eta = 0$ for all individuals.[17]

Panel (a) of Figure 9 plots the excess burden implied by the two sets of estimates as functions of the effective consumption tax rate $\tilde{\tau}$. The solid black line shows the excess burden using our preferred specification. It shows that increasing the marginal tax rate on middle- and high-income individuals results in an economic loss of at least 1 NOK for every additional NOK of tax revenue raised. For $\tilde{\tau} = 0.21$, based on observed expenditure shares, the excess burden is approximately 1.3.

This contrasts sharply with the results obtained from using the conventional specification, as depicted by the dashed grey line. The conventional specification implies a more modest excess burden of 0.23-0.28, depending on the exact value of the effective tax rate $\tilde{\tau}$. We conclude that ignoring income effects and heterogeneity in elasticities can lead to severely downward-biased estimates of excess burden. In our setting, the actual efficiency costs of taxation are four to five times larger than those implied by the conventional specification.

6.3 Revenue-maximizing tax rate

We now turn to deriving and calculating the revenue-maximizing top-income tax rate.

Deriving the revenue-maximizing tax rate. To derive the revenue-maximizing tax rate, it is useful to start by considering the tax revenue collected from individuals with earnings

[17] To see why, note that under homogeneous ETIs and no income effects, Corollary 2 and Proposition 3 together imply that the ETI estimand with $h(Z) = Z$ recovers the compensated elasticity. Hence $\beta = 0.16 = \zeta = \varepsilon^c$.

(a) Excess burden. (b) Revenue-maximizing tax rate.

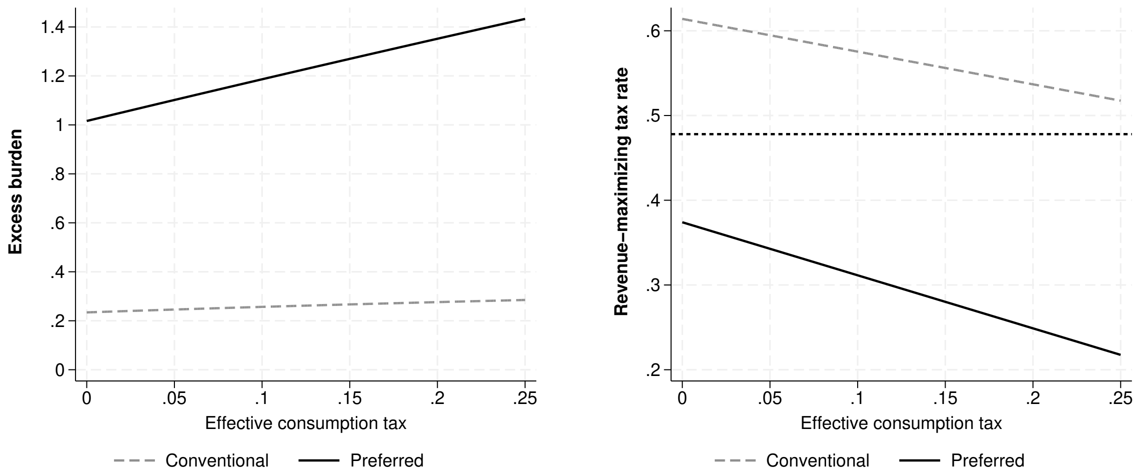


Figure 9: Top-income tax rates and excess burden.

Notes: Panel (a) plots the excess burden of taxation as a function of $\tilde{\tau}$ using our preferred estimates (the solid black line) and using the estimates one would obtain from using the common ETI estimand, ignoring income effects, and assuming constant labor supply elasticities (dashed grey line). Panel (b) plots the revenue-maximizing top-income tax rates as a function of $\tilde{\tau}$ using the same two sets of estimates. The dotted line shows the actual top-income tax rate in Norway after the 2006 reform.

above $\bar{Y}$,

$$\mathbb{E}[TR \mid Y \geq \bar{Y}] = \overbrace{\sum_{k=1}^{K} \tau_k p_k \, \mathbb{E}[C_k \mid Y \geq \bar{Y}]}^{\text{consumption tax}} + \overbrace{\tau \, \mathbb{E}[Y - \bar{Y} \mid Y \geq \bar{Y}] + T(\bar{Y})}^{\text{income tax}} + \underbrace{\tau_w \, \mathbb{E}[Y \mid Y \geq \bar{Y}]}_{\text{payroll tax}}, \tag{32}$$

where the expectations are taken across individuals in the top-income tax bracket.

We characterize the revenue-maximizing top-income tax rate by taking the derivative of equation (32) with respect to the top-income tax rate τ and setting it equal to zero. As we show in Appendix D.2, this first-order condition implies that the following equation is satisfied,

$$\tau_{\text{TOP}} = \frac{\alpha - 1 - (\tilde{\tau} + \tau_w + \tau_w \tilde{\tau})\,(\alpha \bar{\varepsilon}^u - \bar{\eta})}{\alpha - 1 + (\alpha \bar{\varepsilon}^u - \bar{\eta})}, \tag{33}$$

with

$$\bar{\varepsilon}^u \equiv \frac{\mathbb{E}[Y^* \varepsilon^u \mid Y^u \geq \bar{Y}]}{\mathbb{E}[Y^u \mid Y^u \geq \bar{Y}],} \qquad \bar{\eta} \equiv \mathbb{E}[\eta \mid Y^u \geq \bar{Y}], \qquad \alpha \equiv \frac{\mathbb{E}[Y^u \mid Y^u \geq \bar{Y}]}{\bar{Y}}, \tag{34}$$

where $Y^u, \varepsilon^u, \eta, \alpha$, and $\tilde{\tau}$ are evaluated at the revenue-maximizing tax system.

Calculating the revenue-maximizing top-income tax rate. We consider the tax rate that maximizes tax revenue on incomes above $\bar{Y} = 425{,}000$ NOK. Following Saez (2001) and Saez and Stantcheva (2016), we assume that the weighted elasticities $\bar{\varepsilon}^u$, $\bar{\eta}$, and the Pareto parameter α are unaffected by the top tax rate. Using the estimates and data in Section 5, we obtain $\alpha = 1.46$, $\bar{\varepsilon}^u = 0.27$, and $\bar{\eta} = -0.15$. As above, we refer to this set of (weighted-average) elasticities as our preferred specification. For comparison, we also calculate excess burden under the conventional specification, obtained by using the common ETI estimand, ignoring income effects, and assuming constant labor supply elasticities. Under this specification, $\bar{\varepsilon}^u = 0.16$ and $\bar{\eta} = 0$.[18]

Panel (b) of Figure 9 plots the revenue-maximizing top-income tax rates implied by these two specifications as functions of the effective consumption tax rate $\tilde{\tau}$. The solid black line shows the results using our preferred specification. The revenue-maximizing top rate is at most 0.38. It declines with the effective consumption tax rate $\tilde{\tau}$, and reaches 0.22 when $\tilde{\tau}$ is at its upper bound. For $\tilde{\tau} = 0.21$, obtained using observed expenditure shares, the revenue-maximizing tax rate is around 0.25. The gray dashed line uses the elasticities implied by the conventional specification. These estimates increase the revenue-maximizing rates by more than 20 percentage points.

Interestingly, our estimates imply that the actual top-income tax rate after the 2006 reform, illustrated by the dotted black line, exceeds the revenue-maximizing level for any effective consumption tax rate $\tilde{\tau}$. This means that the government could increase revenue by *reducing* top-income tax rates.

7 Implications for Frisch elasticities and EIS

We now consider the implications of our findings for the Frisch elasticity and the elasticity of intertemporal substitution (EIS). We first show how to derive the EIS and Frisch elasticities from the compensated elasticity and income effect estimates in Section 5. We then use these elasticities to consider the efficiency cost of a temporary change in the marginal tax rate.

7.1 Deriving the EIS and Frisch elasticity from income and substitution effects

We consider the textbook intertemporal problem where individuals choose consumption and earnings to maximize the discounted sum of per-period utility subject to the intertemporal budget constraint in equation (24). Following Heathcote et al. (2014), the per-period utility

[18] Under homogeneous ETIs and no income effects, Corollary 2 and Proposition 3 imply that the ETI estimand with $h(Z) = Z$ recovers the compensated elasticity. Hence $\beta = 0.16 = \zeta = \varepsilon^c = \varepsilon^u = \bar{\varepsilon}^u$.

function is given by:

$$U(C,Y) = \frac{C^{1-1/\rho}}{1-1/\rho} - \frac{\alpha}{1+1/\varepsilon^f}\left(\frac{Y}{\alpha}\right)^{1+1/\varepsilon^f},$$

which is a special case of the utility function we considered in Section 4. The EIS is then given by ρ, and the Frisch elasticity is equal to ε^f.

It follows from Keane (2011) that ε^f and ρ are linked to the compensated elasticity and income effect by the following relationships:

$$\varepsilon^f = \frac{\varepsilon^c}{1+\eta}, \qquad \rho = -(1-\tau)\frac{Y_t}{C_t}\frac{\varepsilon^c}{\eta}, \tag{35}$$

where τ is the marginal tax rate and the compensated elasticity ε^c and the income effect η is defined as in Equation (16) in Section 4. These expressions show how our estimates of compensated elasticities and income effects can be used to recover the EIS and the Frisch elasticity. At each level of initial income X, the average EIS and Frisch elasticity are given by:

$$\mathbb{E}[\varepsilon^f \mid X = x] = \frac{\varepsilon^c(x)}{1+\eta(x)},$$
$$\mathbb{E}[\rho \mid X = x] = -\mathbb{E}\left[(1-T'(Y_t))\frac{Y_t}{C_t} \mid X = x\right] \times \frac{\varepsilon^c(x)}{\eta(x)},$$

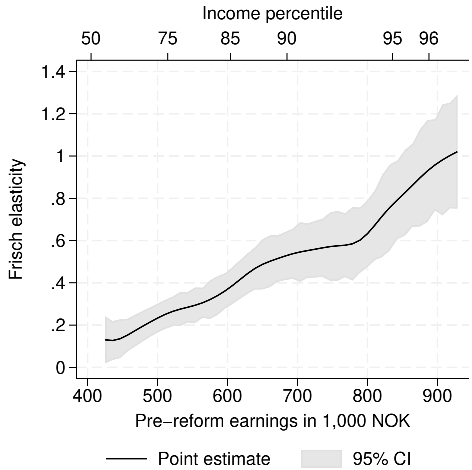
where each component of the RHS can be estimated in the data.

7.2 Estimates of the EIS and Frisch elasticities

Using our estimated compensated elasticities and income effects from Section 5, we find an average Frisch elasticity of 0.4. This finding is close to the average point estimate of 0.5 reported in the meta-analysis of Elminejad et al. (2023). However, as shown in Panel (a) of Figure 10, this average masks considerable heterogeneity. The Frisch elasticity equals 0.15 for incomes around 400,000 (median). It increases steadily to around 0.5 for incomes close to 700,000 (90th percentile), and equals 0.8 for incomes around 850,000 (95th percentile).

In Panel (b) of Figure 10, we report the IES estimates across the income distribution. On average, the EIS is about 1.74. This finding aligns closely with Holm et al.'s (2024) constant effect estimates of the EIS in Norway. However, there is systematic heterogeneity in the EIS across the income distribution. The EIS equals 0.5 for individuals with an income of around 400,000 (median), increases steadily to around 2.5 for incomes close to 700,000 (90th percentile), and is about 4 for incomes around 850,000 (95th percentile).

(a) Frisch elasticity



(b) Elasticity of intertemporal substitution

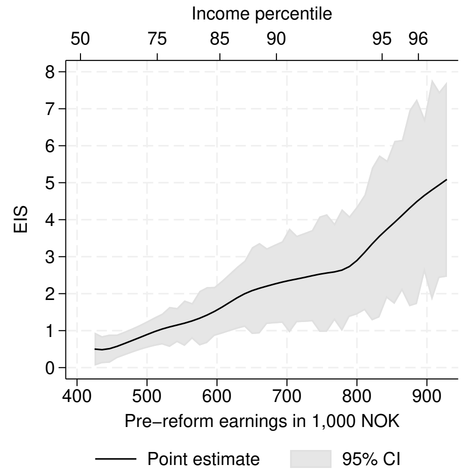


Figure 10: Frisch elasticities and EIS.

Notes: The figure plots point estimates of the Frisch elasticities (Panel (a)) and consumption-weighted EIS (Panel (b)) over the income distribution. 95 percent confidence intervals are shown, with standard errors obtained using the delta method, assuming that the lottery-based income effects $\eta(x)$ and local ETIs $\beta(x)$ are uncorrelated.

7.3 The cost of temporary changes in labor taxation

To illustrate possible implications of the above estimates, we consider the economic cost of a one-period increase in taxes. At time t, the government announces it will temporarily increase the marginal tax rate on labor income by $d\tau$. In period $t+1$, the tax increase is reversed. We measure the efficiency cost of this tax increase as the willingness to pay to avoid the tax change net of the revenue effects. In our model, the willingness to pay is approximately equal to $Y_t d\tau$, while the revenue effect is

$$dTR \approx Y_t d\tau - (\tau + \tau_w) \frac{\varepsilon^f Y_t d\tau}{1-\tau}.$$

We can express the average efficiency cost of the temporary tax conditional on pre-period income $X = x$ as

$$dDWL^f(x) \approx \mathbb{E}\left[\frac{T'(Y_t) + \tau_w}{1 - T'(Y_t)} \mid X = x\right] \varepsilon^f(x) d\tau,$$

where we divide by earnings Y_t before taking the expectation to obtain the efficiency cost as a percentage of pre-tax income.

Figure 11 plots the efficiency cost of a 10 percentage point increase in τ across the distribution of income. The solid line uses the Frisch elasticities reported in Figure 10. It suggests the temporary tax change has an efficiency cost that increases steadily with income. It equals

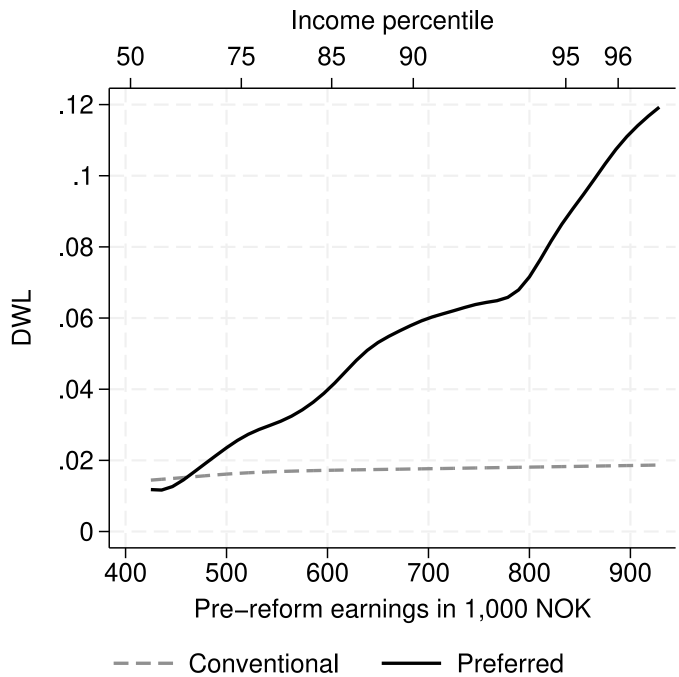


Figure 11: DWL of temporary changes in τ

Notes: The figure plots the implied change in deadweight loss from a one-period ten percentage point increase in the marginal tax rate on labor income, measured in percent of pre-tax earnings.

1.5 percent of the pre-tax income for individuals with an income of around 400,000 (median), increases steadily to around 6 percent for incomes close to 700,000 (90th percentile), and reaches 9 percent for incomes around 850,000 (95th percentile).

We contrast these findings with the conclusions one would obtain from using the common ETI estimand, ignoring income effects, and assuming constant labor supply elasticities. It follows from equation (35) that the Frisch elasticity equals the (un)compensated elasticities with no income effects. The dashed line in Figure 11 plots the implied deadweight loss. For incomes above 500,000 (75th percentile), it grossly understates the true deadweight loss of the tax.

8 Conclusions

This paper examined when a commonly applied class of ETI estimands can be used to learn about individuals' ETI parameters and their (un)compensated elasticities of labor supply. We provided necessary and sufficient conditions for these estimands to be given a causal interpretation as a positively weighted average of heterogeneous ETI parameters. We then applied these results to empirically analyze a reform of the Norwegian tax system that reduced the marginal tax rates on middle and high incomes. The estimated ETI parameters increase steadily with income, meaning high-income individuals are more responsive to tax changes than middle-income individuals.

Next, we showed how (un)compensated elasticities of labor supply can be bounded directly from the ETI estimands, or point identified by combining these estimands with external estimates of income effects. The bounds suggest the compensated and uncompensated elasticities of high-income individuals are at least 0.4 and 0.2, respectively. By comparison, the compensated elasticities of middle-income individuals are bounded between 0.05 and 0.5, while their uncompensated elasticities are close to zero or negative.

We moved from partial to point identification by combining the estimated ETI parameters with estimates of income effects from lottery winnings. The resulting point estimates suggested an (un)compensated elasticity of 0.15 (0.0) for middle-income individuals. The (un)compensated elasticity estimates increase steadily with income to around 0.5 (0.35) for high-income individuals. These findings imply a substantial excess burden of taxation, and that reducing the top-income tax rate would increase tax revenue. Under separable utility, our findings are also informative about how the elasticity of intertemporal substitution and the Frisch elasticity vary across the income distribution.

References

Aaberge, R., & Colombino, U. (2018). Structural labour supply models and microsimulation. *IJM*, *11*(1), 162–197.

Altonji, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy*, *94*(3, Part 2), S176–S215.

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica*, *66*(2), 249–288.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, *90*(430), 431–442.

Arellano, M., & Meghir, C. (1992). Female labour supply and on-the-job search: An empirical model estimated using complementary data sets. *The Review of Economic Studies*, *59*(3), 537–559.

Attanasio, O., Levell, P., Low, H., & Sánchez-Marcos, V. (2018). Aggregating elasticities: Intensive and extensive margins of women's labor supply. *Econometrica*, *86*(6), 2049–2082.

Attanasio, O. P., & Weber, G. (1995). Is consumption growth consistent with intertemporal optimization? evidence from the consumer expenditure survey. *Journal of Political Economy*, *103*(6), 1121–1157.

Auerbach, A. J., & Hines, J. R. J. (2002). Taxation and economic efficiency. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of public economics* (pp. 1347–1421, Vol. 3). Elsevier.

Auten, G., & Carroll, R. (1999). The effect of income taxes on household income. *The Review of Economics and Statistics*, *81*(4), 681–693.

Berg, K., & Thoresen, T. O. (2020). Problematic response margins in the estimation of the elasticity of taxable income. *International Tax and Public Finance*, *27*(3), 721–752.

Blandhol, C., Bonney, J., Mogstad, M., & Torgovitsky, A. (2022, January). *When is tsls actually late?* (Working Paper No. 29709). National Bureau of Economic Research.

Blomquist, N. S. (1985). Labour supply in a two-period model: The effect of a nonlinear progressive income tax. *The Review of Economic Studies*, *52*(3), 515–524.

Blundell, R., & Macurdy, T. (1999). Chapter 27 - labor supply: A review of alternative approaches. In O. C. Ashenfelter & D. Card (Eds.). Elsevier.

Blundell, R., & Walker, I. (1986). A life-cycle consistent empirical model of family labour supply using cross-section data. *The Review of Economic Studies*, *53*(4), 539–558.

Browning, M., & Lusardi, A. (1996). Household saving: Micro theories and micro facts. *Journal of Economic Literature*, *34*(4), 1797–1855.

Bulman, G., Cesarini, D., Lindqvist, E., Notowidigdo, M. J., & Östling, R. (2021). The impact of lottery prizes on winners and their neighbors: Evidence from the dutch postcode lottery. *American Economic Review: Insights*, *3*(3), 333–350.

Burns, S. K., & Ziliak, J. P. (2017). Identifying the elasticity of taxable income. *Economic Journal*, *127*(600), 297–329.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2025). Difference-in-differences with a continuous treatment.

Cesarini, D., Lindqvist, E., Notowidigdo, M. J., & Östling, R. (2017). The effect of wealth on individual and household labor supply: Evidence from swedish lotteries. *American Economic Review*, *107*(12), 3917–3946.

Chetty, R., Guren, A., Manoli, D., & Weber, A. (2011). Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins. *American Economic Review: Papers & Proceedings*, *101*(3), 471–475.

Conesa, J. C., Kitao, S., & Krueger, D. (2009). Taxing capital? not a bad idea after all! *American Economic Review*, *99*(1), 25–48.

Diamond, P. A. (1998). Optimal income taxation: An example with a u-shaped pattern of optimal marginal tax rates. *The American Economic Review*, *88*(1), 83–95.

Eika, L., Mogstad, M., & Vestad, O. L. (2020). What can we learn about household consumption expenditure from data on income and assets? *Journal of Public Economics*, *189*, 104163.

Elminejad, A., Havránek, T., & Irsová, Z. (2023). Intertemporal substitution in labor supply: A meta-analysis. *Labour Economics*, *83*, 102404.

Feldstein, M. (1995). The effect of marginal tax rates on taxable income: A panel study of the 1986 tax reform act. *Journal of Political Economy*, *103*(3), 551–572.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics*, *81*(4), 674–680.

Golosov, M., Graber, M., Mogstad, M., & Novgorodsky, D. (2024). How americans respond to idiosyncratic and exogenous changes in household wealth and unearned income. *The Quarterly Journal of Economics*, *139*(2), 1321–1395.

Gruber, J., & Saez, E. (2002). The elasticity of taxable income: Evidence and implications. *Journal of public Economics*, *84*(1), 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)*.

Hall, R. E. (1988). Intertemporal substitution in consumption. *Journal of Political Economy*, *96*(2), 339–357.

Hansen, L. P., & Singleton, K. J. (1983). Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, *91*(2), 249–265.

Harberger, A. (1964). Taxation, resource allocation, and welfare. In *The role of direct and indirect taxes in the federal reserve system* (pp. 25–80). National Bureau of Economic Research, Inc.

Havránek, T. (2015). Measuring intertemporal substitution: The effect of different data. *Journal of the European Economic Association*, *13*(6), 1180–1204.

Heathcote, J., Storesletten, K., & Violante, G. L. (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review*, *104*(7), 2075–2126.

Holm, M. B., Jamilov, R., Jasinski, M., & Nenov, P. T. (2024). *Estimating the elasticity of intertemporal substitution using dividend tax news shocks* (tech. rep. No. 17/2024). Norges Bank Working Paper.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Imbens, G. W., Rubin, D. B., & Sacerdote, B. I. (2001). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American Economic Review*, *91*(4), 778–794.

Keane, M., & Neal, T. (2025). Robust inference for the frisch labor supply elasticity. *Journal of Labor Economics*, *43*(S1), S179–S219.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, *49*(4), 961–1075.

Killingsworth, M. R., & Heckman, J. J. (1986). Female labor supply: A survey. In O. Ashenfelter & R. Layard (Eds.), *Handbook of labor economics* (pp. 103–204, Vol. 1). Elsevier.

Kleven, H. J., & Schultz, E. A. (2014). Estimating taxable income responses using danish tax reforms. *American Economic Journal: Economic Policy*, *6*(4), 271–301.

Kumar, A., & Liang, C.-Y. (2020). Estimating taxable income responses with elasticity heterogeneity. *Journal of Public Economics*, *188*, 104209.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy*, *89*(6), 1059–1085.

MaCurdy, T. E. (1983). A simple scheme for estimating an intertemporal model of labor supply and consumption in the presence of taxes and uncertainty. *International Economic Review*, *24*(2), 265–289.

Mogstad, M., Salvanes, K. G., & Torsvik, G. (2025). Income equality in the nordic countries: Myths, facts, and lessons. *Journal of Economic Literature*, *63*(3), 791–839.

Mogstad, M., & Torgovitsky, A. (2024). Chapter 1 - instrumental variables with unobserved heterogeneity in treatment effects. In C. Dustmann & T. Lemieux (Eds.). Elsevier.

Neisser, C. (2021). The elasticity of taxable income: A meta-regression analysis. *The Economic Journal*, *131*(640), 3365–3391.

Pencavel, J. (1986). Labor supply of men: A survey (O. Ashenfelter & R. Layard, Eds.). *Handbook of Labor Economics*, *1*, 3–102.

Picchio, M., Suetens, S., & van Ours, J. C. (2018). Labor supply effects of winning a lottery. *Economic Journal*, *128*(610), 1700–1729.

Prescott, E. C. (2006). Nobel lecture: The transformation of macroeconomic policy and research. *Journal of Political Economy*, *114*(2), 203–235.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates. *The Review of Economic Studies*, *68*(1), 205–229.

Saez, E., Slemrod, J., & Giertz, S. H. (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of economic literature*, *50*(1), 3–50.

Saez, E., & Stantcheva, S. (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review*, *106*(1), 24–45.

Straub, L., & Werning, I. (2020). Positive long-run capital taxation: Chamley-judd revisited. *American Economic Review*, *110*(1), 86–119.

Thoresen, T. O., & Vattø, T. E. (2015). Validation of the discrete choice labor supply model by methods of the new tax responsiveness literature. *Labour Economics*, *37*, 38–53.

Vattø, T. E. (2020). Estimating the elasticity of taxable income when earnings responses are sluggish [Articles]. *Articles*, *76*(- 4), 329–369.

Weber, C. E. (2014). Toward obtaining a consistent estimate of the elasticity of taxable income using difference-in-differences. *Journal of Public Economics*, *117*, 90–103.

A Proofs

Many of our results in Section 2 build on the following two lemmas, which map the ETI estimand β to the potential outcomes for any h provided f is unrestricted.

Lemma 1. *Suppose Assumptions 1 and 2 are true and let f be unrestricted over the support of X. Then, the ETI estimand equals a weighted average conditional-on-X elasticities of taxable income,*

$$\beta = \sum_{j=k_0}^{K} \omega_j^h \zeta(x_j), \tag{36}$$

where,

$$\omega_j^h \equiv \frac{p_j \phi(x_j)\, \mathbb{E}[D^h \mid G = 1, X = x_j]}{\sum_{l=k_0}^{K} p_l \phi(x_l)\, \mathbb{E}[D^h \mid G = 1, X = x_l]},$$

where $p_j \equiv \mathbb{P}(X = x_j \mid G = 1)$, D^h denotes the predicted residuals resulting from regressing $Gh(Z)$ on G and $f(X;.)$, and

$$\begin{aligned}
\phi(x) &\equiv \mathbb{E}[\text{NTR}(1) - \text{NTR}(0) \mid G = 1, X = x], \\
\zeta(x) &\equiv \mathbb{E}\left[\frac{\text{NTR}(1) - \text{NTR}(0)}{\mathbb{E}[\text{NTR}(1) - \text{NTR}(0) \mid G = 1, X = x]} \times \zeta \mid G = 1, X = x\right].
\end{aligned}$$

Proof of Lemma 1. We start by deriving the expression for the ETI estimand. This requires some additional notation.

For some fixed function h, the Frisch-Waugh-Lovell theorem allows for expressing the ETI estimand as,

$$\beta \equiv \frac{\mathbb{E}\left[\Delta y D^h\right]}{\mathbb{E}\left[\Delta\, \text{NTR}\, D^h\right]}, \tag{37}$$

where

$$D^h \equiv Gh(Z) - \theta_0 G - f(X;\theta) \tag{38}$$

denotes predicted residuals from the linear projection of $Gh(Z)$ on $(G, f(X;.))$ where the coefficients are given as the solution to,

$$(\theta_0, \theta) \equiv \operatorname*{arg\,min}_{\theta_g, \theta_x} \mathbb{E}\left[\left(Gh(Z) - \theta_g G - f(X;\theta_x)\right)^2\right]. \tag{39}$$

Since f is unrestricted, it can be written as:

$$f(x;\theta) = \sum_{k=1}^{K} \theta_k \mathbb{1}[x = x_k], \tag{40}$$

where $\{x_1, ..., x_K\}$ is the support of X.

It is useful to denote by $\rho \equiv y(1) - y(0)$ the earnings response to the tax reform. Then, by recalling equation (5), we get that $y = y(0) + \rho \mathbb{1}[Z \neq 0]G$, which implies that $\Delta y = \Delta y(0) + \rho \mathbb{1}[Z \neq 0]G$. Using equation (8) from Assumption 1, we can write,

$$\Delta y = \lambda^y G + \rho \mathbb{1}[Z \neq 0]G + f_y(X) + u, \tag{41}$$

with $\mathbb{E}[u \mid G, X] = 0$. Plugging (41) into the numerator of the ETI estimand in equation (37) gives,

$$\begin{aligned}
\mathbb{E}[\Delta y D^h] &= \mathbb{E}\left[\left(\lambda^y G + \rho \mathbb{1}[Z \neq 0]G + f_y(X) + u\right) D^h\right], \\
&= \mathbb{E}\left[\rho \mathbb{1}[Z \neq 0] G D^h\right], \\
&= \mathbb{P}(GZ \neq 0)\, \mathbb{E}[\rho D^h \mid GZ \neq 0],
\end{aligned}$$

where the second equality follows from the fact that $\mathbb{E}[u \mid G, X] = 0$ and that the moment conditions that determine D^h since the unrestricted specification of f ensures that $f(\cdot;\theta) = f_y(\cdot)$ for some θ. The third equality follows from the law of total expectations.

Inserting for the linear projection D^h,

$$\mathbb{E}\left[\Delta y D^h\right] = \mathbb{P}(GZ \neq 0)\, \mathbb{E}\left[\rho D^h \mid GZ \neq 0\right],$$

Similar reasoning yields the following expression for the denominator of equation (37),

$$\mathbb{E}\left[\Delta\, \mathrm{NTR}\, D^h\right] = \mathbb{P}(GZ \neq 0)\, \mathbb{E}\left[\phi D^h \mid GZ \neq 0\right].$$

By combining the two terms, we obtain,

$$\begin{aligned}
\beta &= \frac{\mathbb{E}\left[\phi D^h \times \zeta \mid GZ \neq 0\right]}{\mathbb{E}\left[\phi D^h \mid GZ \neq 0\right]}, \\
&= \frac{\mathbb{E}\left[\phi D^h \times \zeta \mid G = 1, X \geq x_{k_0}\right]}{\mathbb{E}\left[\phi D^h \mid G = 1, X \geq x_{k_0}\right]}
\end{aligned}$$

where we have used that $\mathbb{P}(GZ \neq 0)$ cancels and that Assumption 2 implies that $\rho = \phi \times \zeta$. From the definition of the simulated instrument Z, it is clear that X deterministically determines Z. Thus, D^h is a deterministic function of G and X, meaning we can write the realization of D^h for an individual with $G = g$ and $X = x$ as $D^h(g, x)$. We obtain,

$$\begin{aligned}
\beta &= \frac{\sum_{j=k_0}^{K} p_j D^h(1, x_j)\, \mathbb{E}\left[\phi \times \zeta \mid G = 1, X = x_j\right]}{\sum_{j=k_0}^{K} p_j D^h(1, x_j)\, \mathbb{E}\left[\phi \mid G = 1, X = x_j\right]}, \\
&= \frac{\sum_{j=k_0}^{K} p_j D^h(1, x_j)\phi(x_j)\zeta(x_j)}{\sum_{j=k_0}^{K} p_j D^h(1, x_j)\phi(x_j)}, \\
&= \sum_{j=k_0}^{K} \frac{p_j D^h(1, x_j)\phi(x_j)}{\sum_{m=k_0}^{K} p_m D^h(1, x_m)\phi(x_m)} \times \zeta(x_j).
\end{aligned}$$

The result follows from noting that $D^h(1, x_k) = \mathbb{E}[D^h \mid G = 1, X = x_k]$ and using the definition of ω_j^h.
□

Lemma 2. *Suppose Assumptions 1 and 2 are true and let f be flexible. Then,*

$$D^h(1,x_k) = \mathbb{P}(G=0 \mid X=x_k)\Bigg(\mathbb{E}[h(Z) \mid X=x_k] - \frac{\sum_{j=1}^K p_j \, \mathbb{P}(G=0 \mid X=x_j)\, \mathbb{E}[h(Z) \mid X=x_j]}{\sum_{j=1}^K p_j \, \mathbb{P}(G=0 \mid X=x_j)} \Bigg).$$

Proof of Lemma 2. When f is unrestricted, the first-order conditions to the minimization problem in equation (39) yield the following moment conditions

$$\mathbb{E}\left[G\left(Gh(Z) - \theta_0 - \sum_{k=1}^K \theta_k \mathbb{1}[X=x_k] \right) \right] = 0, \tag{42}$$

$$\mathbb{E}\left[\mathbb{1}[X=x_k] \left(Gh(Z) - \theta_0 G - \sum_{j=1}^K \theta_j \mathbb{1}[X=x_j] \right) \right] = 0, \tag{43}$$

for $j = 1, ...K$. From (43), we obtain

$$\mathbb{E}\left[Gh(Z) - \theta_0 G - \theta_k \mid X = x_k \right] = 0.$$

Since Z is a deterministic function of X, this can be rewritten as

$$\theta_k = \mathbb{P}(G=1 \mid X=x_k)\left(\mathbb{E}[h(Z) \mid X=x_k] - \theta_0 \right). \tag{44}$$

Next, we use equation (42) which implies that,

$$\mathbb{E}[h(Z) \mid G=1] = \theta_0 + \sum_{k=1}^K p_k \theta_k, \tag{45}$$

where $p_k \equiv \mathbb{P}(X = x_k \mid G = 1)$. Substituting the expression of θ_k from equation (44) into the expression above yields,

$$\mathbb{E}[h(Z) \mid G=1] = \theta_0 + \sum_{k=1}^K p_k \, \mathbb{P}(G=1 \mid X=x_k)\left(\mathbb{E}[h(Z) \mid X=x_k] - \theta_0 \right),$$

and solving for θ_0 gives,

$$\begin{aligned}
\theta_0 &= \frac{\mathbb{E}[h(Z) \mid G=1] - \sum_{k=1}^K p_k \, \mathbb{P}(G=1 \mid X=x_k)\, \mathbb{E}[h(Z) \mid X=x_k]}{1 - \sum_{k=1}^K p_k \, \mathbb{P}(G=1 \mid X=x_k)}, \\
&= \frac{\sum_{k=1}^K p_k \left(1 - \mathbb{P}(G=1 \mid X=x_k)\right) \mathbb{E}[h(Z) \mid X=x_k]}{1 - \sum_{k=1}^K p_k \, \mathbb{P}(G=1 \mid X=x_k)}, \\
&= \frac{\sum_{k=1}^K p_k \left(1 - \mathbb{P}(G=1 \mid X=x_k)\right) \mathbb{E}[h(Z) \mid X=x_k]}{\sum_{k=1}^K p_k (1 - \mathbb{P}(G=1 \mid X=x_k))}.
\end{aligned}$$

$D^h(1, x_k)$ can then be rewritten as,

$$
\begin{aligned}
D^h(1, x_k) &= \mathbb{E}[h(Z) \mid X = x_k] - \theta_0 - \mathbb{P}(G = 1 \mid X = x_k) \left(\mathbb{E}[h(Z) \mid X = x_k] - \theta_0\right), \\
&= \left(1 - \mathbb{P}(G = 1 \mid X = x_k)\right) \left(\mathbb{E}[h(Z) \mid X = x_k] - \theta_0\right), \\
&= \left(1 - \mathbb{P}(G = 1 \mid X = x_k)\right) \\
&\quad \times \left(\mathbb{E}[h(Z) \mid X = x_k] - \frac{\sum_{j=1}^{K} p_j \left(1 - \mathbb{P}(G = 1 \mid X = x_j)\right) \mathbb{E}[h(Z) \mid X = x_j]}{\sum_{j=1}^{K} p_j (1 - \mathbb{P}(G = 1 \mid X = x_j))}\right), \\
&= \mathbb{P}(G = 0 \mid X = x_k) \\
&\quad \times \left(\mathbb{E}[h(Z) \mid X = x_k] - \frac{\sum_{j=1}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j) \, \mathbb{E}[h(Z) \mid X = x_j]}{\sum_{j=1}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}\right),
\end{aligned}
$$

where the first equality uses equation (44) and the second substituties in the expression for θ_0.

□

Proof of Proposition 1. We start by proving the if-part of the statement.

If $h(Z)$ is binary and f is flexible then β is causal: Since $h(Z)$ is binary we can write $h(Z) = h_0 \mathbb{1}[Z = 0] + h_1 \mathbb{1}[Z \neq 0]$. Lemma 2 then implies that

$$
\begin{aligned}
D^h(1, x_k) =& \, \mathbb{P}(G = 0 \mid X = x_k) \\
&\times \left(h_1 - \frac{h_0 \sum_{j=1}^{k_0 - 1} p_j \, \mathbb{P}(G = 0 \mid X = x_j) + h_1 \sum_{j=k_0}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}{\sum_{j=1}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}\right), \\
=& \, \mathbb{P}(G = 0 \mid X = x_k) \\
&\times \left(h_0 + (h_1 - h_0) - \frac{(h_1 - h_0) \sum_{j=k_0}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}{\sum_{j=1}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)} - h_0\right), \\
=& \, \mathbb{P}(G = 0 \mid X = x_k) \times (h_1 - h_0) \times \left(1 - \frac{\sum_{j=k_0}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}{\sum_{j=1}^{K} p_j \, \mathbb{P}(G = 0 \mid X = x_j)}\right).
\end{aligned}
$$

for any $x_k \in \{x_{k_0}, ..., x_K\}$. According to Lemma 1, $D^h(1, x_k)$ appears both in the numerator and denominator of the expression for β. Since the last two components of $D^h(1, x_k)$ do not vary with x_k, they cancel out, and we obtain:

$$
\beta = \sum_{j=k_0}^{K} \frac{\mathbb{P}(X = x_j \mid G = 1) \phi(x_j) \, \mathbb{P}(G = 0 \mid X = x_j)}{\sum_{l=k_0}^{K} \mathbb{P}(X = x_l \mid G = 1) \phi(x_l) \, \mathbb{P}(G = 0 \mid X = x_l)} \zeta(x_j). \tag{46}
$$

By Bayes law, $\mathbb{P}(X = x_j \mid G = 1) = \mathbb{P}(G = 1 \mid X = x_j)\,\mathbb{P}(X = x_j)$, so we obtain,

$$\beta = \sum_{j=k_0}^{K} \frac{\mathbb{P}(X = x_j)\,\mathbb{P}(G = 1 \mid X = x_j)\,\mathbb{P}(G = 0 \mid X = x_j)\phi(x_j)}{\sum_{l=k_0}^{K} \mathbb{P}(X = x_l)\,\mathbb{P}(G = 1 \mid X = x_l)\,\mathbb{P}(G = 0 \mid X = x_l)\phi(x_l)} \zeta(x_j), \tag{47}$$

$$= \sum_{j=k_0}^{K} \frac{\mathbb{P}(X = x_j) Var(G \mid X = x_j)\phi(x_j)}{\sum_{l=k_0}^{K} \mathbb{P}(X = x_l) Var(G \mid X = x_l)\phi(x_l)} \zeta(x_j), \tag{48}$$

where the second equality uses that G follows a Bernoulli distribution, implying that $\mathbb{P}(G = 1 \mid X = x_j)\,\mathbb{P}(G = 0 \mid X = x_j) = Var(G \mid X = x_j)$.

Assumption 3 ensures that $\zeta(x_k)$ is a positively weighted average ζ and that the sign of $\phi(x_k)$ does not change with k. Since probabilities and variances are non-negative, it follows that the ETI estimand recovers a positively weighted average of individual elasticities.

If $h(Z)$ is not binary and f is flexible then β is not causal: For β to be causal, it must be the case that the sign of,

$$\left(\mathbb{E}[h(Z) \mid X = x_k] - \frac{\sum_{j=1}^{K} p_j\, \mathbb{P}(G = 0 \mid X = x_j)\,\mathbb{E}[h(Z) \mid X = x_j]}{\sum_{j=1}^{K} p_j\, \mathbb{P}(G = 0 \mid X = x_j)} \right) \tag{49}$$

does not change with $k_0 \geq k \leq K$ for any distribution of X. We prove that it is always possible to construct a distribution of X where these weights change sign in the case when $h(Z)$ takes on more than two values. Let,

$$h(Z) = \begin{cases} h_0 & \text{if } X < \bar{Y}, \\ \sum_{k=k_0}^{K} \mathbb{1}[X = x_k] h_k & \text{if } X = \bar{Y}. \end{cases} \tag{50}$$

where $(h_0, h_{k_0}, ..., h_K)$ is the support of $h(Z)$. Using this, we can write the weights as

$$\sum_{k=k_0}^{K} h_k \mathbb{1}[X = x_k] - \frac{\sum_{j=1}^{K} \mathbb{P}(X = x_j) Var(G \mid X = x_j) h_k}{\sum_{j=1}^{K} \mathbb{P}(X = x_j) Var(G \mid X = x_j)}.$$

The last term is a convex combination of the points in the support of $h(Z)$. This means that, depending on the distribution of X, it can take on any value between the smallest and largest value of h_k. Thus, it is always possible to construct a distribution of X such that the convex combination is between any two values $h_{k'}, h_{k''}$ with $k', k'' \geq k_0$, meaning negative weights will emerge.

If $h(Z)$ is binary and f is not flexible then β is not causal: Suppose that ζ is constant across individuals and consider the IV regression model:

$$\Delta y = \alpha_0^y G + \beta \Delta\,\text{NTR} + f(X; \alpha^y) + u^y, \tag{51}$$

$$\Delta\,\text{NTR} = \alpha_0^{\text{NTR}} G + \alpha G \mathbb{1}[Z \neq 0] + f(X; \alpha^{\text{NTR}}) + u^{\text{NTR}}. \tag{52}$$

Assumptions 1 and 2 then imply that

$$\Delta y = \lambda^y G + f(X;\alpha^y) + \zeta \Delta \text{ NTR} + u^y. \tag{53}$$

with $u^y \equiv f_y(X) - f(X;\theta) + \Delta y(0) - \mathbb{E}[\Delta y(0) \mid X, G]$. The standard exogeneity assumption requires that $Cov(u^y, G\mathbb{1}[Z \neq 0]) = 0$. Note that,

$$\begin{aligned}
Cov(u^y, G\mathbb{1}[Z \neq 0]) &= Cov(f_y(X) - f(X;\theta) + \Delta y(0) - \mathbb{E}[\Delta y(0) \mid X, G], G\mathbb{1}[Z \neq 0]),\\
&= Cov(f_y(X) - f(X;\theta), G\mathbb{1}[Z \neq 0]),\\
&= \mathbb{P}(GZ \neq 0)\Big(\mathbb{E}[f_y(X) - f(X;\theta) \mid GZ \neq 0]\\
&\qquad\qquad - \mathbb{E}[f_y(X) - f(X;\theta)]\Big).
\end{aligned}$$

Unless f is unrestricted so that $f_y(X) = f(X;\theta)$ for all X, $\mathbb{E}[f_y(X) - f(X;\theta) \mid GZ \neq 0] \neq \mathbb{E}[f_y(X) - f(X;\theta)]$ meaning that exogeneity generally fails. □

Proof of Corollary 1. From the proof of Proposition 1, we have that

$$\beta = \sum_{j=k_0}^{K} \frac{\mathbb{P}(X = x_j) Var(G \mid X = x_j)\phi(x_j)}{\sum_{l=k_0}^{K} \mathbb{P}(X = x_l) Var(G \mid X = x_l)\phi(x_l)} \zeta(x_j). \tag{54}$$

Note that by the law of iterated expectaions we obtain

$$\begin{aligned}
\zeta(x_k) &= \sum_{j=1}^{J} \frac{\mathbb{P}(\phi = \phi_j \mid G = 1, X = x_k)\phi_j}{\sum_{l=1}^{J} \mathbb{P}(\phi = \phi_l \mid G = 1, X = x_k)\phi_l} \mathbb{E}[\zeta \mid G = 1, X = x_k, \phi = \phi_j],\\
\phi(x_k) &= \sum_{j=1}^{J} \mathbb{P}(\phi = \phi_j \mid G = 1, X = x_k)\phi_j.
\end{aligned}$$

Letting $\zeta_{k,j} \equiv \mathbb{E}[\zeta \mid G=1, X=x_k, \phi=\phi_j]$ and substituting these equations into (54), we obtain

$$
\begin{aligned}
\beta &= \sum_{j=k_0}^{K} \frac{\mathbb{P}(X=x_j)Var(G \mid X=x_j)\phi(x_j)}{\sum_{l=k_0}^{K}\mathbb{P}(X=x_l)Var(G \mid X=x_l)\phi(x_l)} \frac{\sum_{j=1}^{J}\mathbb{P}(\phi=\phi_j \mid G=1, X=x_k)\phi_j}{\phi(x_k)} \zeta_{k,l}, \\
&= \frac{\sum_{j=k_0}^{K}\sum_{j=1}^{J}\mathbb{P}(X=x_j)\,\mathbb{P}(\phi=\phi_j \mid G=1, X=x_k)Var(G \mid X=x_j)\phi_j\zeta_{k,l}}{\sum_{l=k_0}^{K}\mathbb{P}(X=x_l)Var(G \mid X=x_l)\phi(x_l)}, \\
&= \frac{\sum_{j=k_0}^{K}\sum_{j=1}^{J}\mathbb{P}(X=x_j, \phi=\phi_j \mid G=1)Var(G \mid X=x_j)\phi_j\zeta_{k,l}}{\sum_{l=k_0}^{K}\mathbb{P}(X=x_l)Var(G \mid X=x_l)\phi(x_l)}, \\
&= \frac{\sum_{j=k_0}^{K}\sum_{j=1}^{J}\mathbb{P}(X=x_j, \phi=\phi_j \mid G=1)Var(G \mid X=x_j)\phi_j\zeta_{k,l}}{\sum_{l=k_0}^{K}\mathbb{P}(X=x_l)Var(G \mid X=x_l)\sum_{j=1}^{J}\mathbb{P}(\phi=\phi_j \mid G=1, X=x_l)\phi_j}, \\
&= \frac{\sum_{j=k_0}^{K}\sum_{j=1}^{J}\mathbb{P}(X=x_j, \phi=\phi_j \mid G=1)Var(G \mid X=x_j)\phi_j\zeta_{k,l}}{\sum_{l=k_0}^{K}\sum_{j=1}^{J}Var(G \mid X=x_l)\,\mathbb{P}(X=x_l, \phi=\phi_j \mid G=1, X=x_l)\phi_j}, \\
&= \sum_{j=k_0}^{K}\sum_{j=1}^{J} \frac{\mathbb{P}(X=x_j, \phi=\phi_j \mid G=1)Var(G \mid X=x_j)\phi_j}{\sum_{l=k_0}^{K}\sum_{j=1}^{J}Var(G \mid X=x_l)\,\mathbb{P}(X=x_l, \phi=\phi_j \mid G=1, X=x_l)\phi_j} \times \zeta_{k,l}.
\end{aligned}
$$

□

Proof of Corollary 2. When ζ is constant across individuals, then $\zeta(x_k) = \zeta(x_l) = \zeta$ for any k, l. Thus, by Lemma 1:

$$
\begin{aligned}
\beta &= \sum_{j=k_0}^{K} \frac{p_j D^h(1, x_j)\phi_j}{\sum_{m=k_0}^{K} p_m D^h(1, x_m)\phi_m} \times \zeta, \\
&= \zeta \frac{\sum_{j=k_0}^{K} p_j D^h(1, x_j)\phi_j}{\sum_{m=k_0}^{K} p_m D^h(1, x_m)\phi_m} \\
&= \zeta.
\end{aligned}
$$

□

Proof of Proposition 2. Start by noting that equation (5) and the definition of k_0 imply

$$
\mathbb{E}[\Delta y \mid G, X] = \begin{cases} \mathbb{E}[\Delta y(0) + \rho \mid G, X] & \text{if } G=1 \text{ and } X \geq x_{k_0}, \\ \mathbb{E}[\Delta y(0) \mid G, X] & \text{otherwise.} \end{cases}
$$

Thus, under assumption 1, the regression of Δy on G and dummies for each value of X on the sub-sample with $GZ = 0$ recovers

$$
\mathbb{E}[\Delta y(0) \mid G=g, X=x] = \hat{\lambda}_0^y g + \sum_{k=1}^{K} \hat{\lambda}_k^y \mathbb{1}[x = x_k],
$$

for each x in the support of X. Thus, for any $x \geq x_{k_0}$, the numerator of $\beta(x)$ equals,

$$\begin{aligned}
&\mathbb{E}[\Delta y \mid G=1, X=x] - \left(\hat{\lambda}_0^y g + \sum_{k=k_0}^{K} \hat{\lambda}_k^y \mathbb{1}[x = x_k]\right), \\
=&\mathbb{E}[\Delta y(0) + \rho \mid G=1, X=x] - \mathbb{E}[\Delta y(0) \mid G=1, X=x], \\
=&\mathbb{E}[\phi \times \zeta \mid G=1, X=x], \\
=&\phi(x) \times \zeta(x),
\end{aligned}$$

where the second equality follows from Assumption 2. Corresponding arguments show that the denominator of the $\beta(x)$ equals,

$$\mathbb{E}[\Delta \,\text{NTR} \mid G=1, X=x] - \left(\hat{\lambda}_0^{\text{NTR}} g + \sum_{k=k_0}^{K} \hat{\lambda}_k^{\text{NTR}} \mathbb{1}[x = x_k]\right) = \phi(x).$$

Thus, the local ETI estimand can be written as,

$$\begin{aligned}
\beta(x) &= \frac{\phi(x)\zeta(x)}{\phi(x)} = \zeta(x) \\
&= \sum_{j=1}^{J} \frac{\mathbb{P}(X=x, \phi=\phi_j \mid G=1)\phi_j}{\sum_{l=1}^{J} \mathbb{P}(X=x, \phi=\phi_l \mid G=1)\phi_l} \times \mathbb{E}[\zeta \mid G=1, X=x, \phi=\phi_j],
\end{aligned}$$

where the second line uses the law of iterated expectations, completing the proof. □

Proof of Proposition 3. The result follows from the proof of Proposition 4. Specifically, by noting that with no income effects ($\eta = 0$),

$$\log Y(\tau, R) \approx y_t + \varepsilon^u \log\left(\frac{1-\tau}{1-\tau_t}\right) + \frac{\eta(R - R_t)}{(1-\tau_t)Y_t} = y_t + \varepsilon^c \log\left(\frac{1-\tau}{1-\tau_t}\right).$$

Thus,

$$\log Y(\tau(d), R(d)) = y_t + \varepsilon^c \log\left(\frac{1-\tau(d)}{1-\tau_t}\right) = \zeta \log\left(\frac{1-\tau(d)}{1-\tau_t}\right) + \nu(d),$$

meaning that $\zeta = \varepsilon^c$ and $y_t = \nu(0) = \nu(1)$. □

Proof of Proposition 4. We start by deriving the first-order approximation to the (log-)earnings function $\log Y(\tau, R)$ with respect to changes in $\log(1-\tau)$ and R around the observed virtual income tax system (τ_t, R_t). Start by noting that $\log Y(\tau, R) = \log Y\left(1 - \exp(\log(1-\tau)), R\right)$. Taking the derivative with respect to $\log(1-\tau)$ then gives,

$$\frac{\partial \log Y}{\partial \log(1-\tau)} = \frac{\partial Y}{\partial 1-\tau} \frac{1-\tau}{Y} = \varepsilon^u.$$

Taking the derivative with respect to R,

$$\frac{\partial \log Y}{\partial R} = \frac{1}{Y}\frac{\partial Y}{\partial R} = \frac{\eta}{(1-\tau)Y}.$$

This means the first-order approximation to $\log Y(\tau, R)$ around τ_t, R_t equals,

$$\log Y(\tau, R) \approx y_t + \varepsilon^u \log\left(\frac{1-\tau}{1-\tau_t}\right) + \frac{\eta(R-R_t)}{(1-\tau_t)Y_t}.$$

Substituting in the potential marginal tax rates and virtual incomes defined in equation (18) gives,

$$\begin{aligned} y(\tau(1), R(1)) - y(\tau(0), R(0)) &= \varepsilon^u \log\left(\frac{1-\tau(1)}{1-\tau(0)}\right) + \frac{\eta(R(1)-R(0))}{(1-\tau_t)Y_t}, \\ &= \varepsilon^c \phi + \eta\left(\phi + \frac{R(1)-R(0)}{(1-\tau_t)Y_t}\right), \end{aligned}$$

where the second equality uses the Slutsky equation and the definition of ϕ. Now, consider a treated individual (i.e., with $GZ \neq 0$). Focusing on the last term, we can write,

$$\begin{aligned} \phi + \frac{R(1)-R(0)}{(1-\tau_t)Y_t} &= \log\left(\frac{1-\tau(1)}{1-\tau(0)}\right) + \frac{\tau(1)Y(1) - T_1(Y(1)) - (\tau(0)Y(0) - T_0(Y(0)))}{(1-\tau_t)Y_t}, \\ &\approx -\frac{\tau(1)-\tau(0)}{1-\tau(1)} + \frac{\tau(1)Y(1) - T_1(Y(1)) - (\tau(0)Y(1) - T_0(Y(1)))}{(1-\tau_t)Y_t}, \\ &= -\frac{\tau(1)-\tau(0)}{1-\tau(1)} + \frac{(\tau(1)-\tau(0))Y(1)}{(1-\tau(1))Y(1)} + \frac{-T_1(Y(1)) + T_0(Y(1))}{(1-\tau_t)Y_t}, \\ &= -\frac{T_1(Y_t) - T_0(Y_t)}{(1-\tau_t)Y_t}, \end{aligned}$$

where the first equality uses the definitions of $\tau(d)$ and $R(d)$. The second eqality is obtained from taking a first-order approximation of the the log term around $\tau(1) = \tau_t$ and using the no bracket switching assumption, while the third and forth lines obtain from rearranging and exploiting that $Y_t = Y(1)$ and $\tau_t = \tau(1)$ provided $GZ \neq 0$.

From the proof of Proposition 2, we obtain that,

$$\begin{aligned} \beta(x) &= \frac{\mathbb{E}\left[y(1) - y(0) \mid G = 1, X = x\right]}{\phi(x)}, \\ &= \frac{\mathbb{E}\left[\varepsilon^c \phi + \eta\left(\phi + \frac{R(1)-R(0)}{(1-\tau_t)Y_t}\right) \mid G = 1, X = x\right]}{\phi(x)}, \\ &= \varepsilon^c(x) + \eta(x)\frac{\mathbb{E}\left[\left(\phi + \frac{R(1)-R(0)}{(1-\tau_t)Y_t}\right) \mid G = 1, X = x\right]}{\phi(x)}, \\ &= \varepsilon^c(x) - \eta(x)\frac{\mathbb{E}\left[\frac{T_1(Y_t)-T_0(Y_t)}{(1-\tau_t)Y_t} \mid G = 1, X = x\right]}{\phi(x)}. \end{aligned}$$

Under Assumption 1, we have that

$$\phi(x) = \mathbb{E}[\Delta\,\mathrm{NTR} \mid G = 1, X = x] - \left(\hat{\lambda}_0^{\mathrm{NTR}} g + \sum_{k=k_0}^{K} \hat{\lambda}_k^{\mathrm{NTR}} \mathbb{1}[x = x_k]\right).$$

Re-arranging gives the first part of the result:

$$\varepsilon^c(x) = \beta(x) + \eta(x) \frac{\mathbb{E}\left[\frac{T_1(Y_t) - T_0(Y_t)}{(1-\tau_t)Y_t} \mid G = 1, X = x\right]}{\mathbb{E}[\Delta\,\mathrm{NTR} \mid G = 1, X = x] - \left(\hat{\lambda}_0^{\mathrm{NTR}} g + \sum_{k=k_0}^{K} \hat{\lambda}_k^{\mathrm{NTR}} \mathbb{1}[x = x_k]\right)}.$$

The second part of the result follows by adding $\eta(x)$ to both sides of the equation and using the Slutsky equation: $\varepsilon^u(x) = \varepsilon^c(x) + \eta(x)$.

□

B Appendix figures and tables

Table 4: Summary statistics for ETI sample

	Full sample	Estimation sample	
		Reform cohorts	Placebo cohorts
		Demographics	
Age	40.70	40.86	40.66
Education	12.14	12.31	12.29
Male	57.20	62.80	64.80
Married	61.60	57.70	61.30
		Income and taxes	
Taxable income X	443	470	468
Income taxes $T(X)$	130	131	139
Marginal tax rate $T'(X)$	0.414	0.424	0.430
Observations	5,702,759	2,333,992	1,677,302
Individuals	1,346,358	961,832	796,533

Notes: This table reports summary statistics for the sample used in the estimation of the elasticity of taxable income (ETI). Monetary values are consumer-price-index adjusted and reported in 1,000 2018 NOKs (8.13 NOK/USD), and binary outcomes are reported in percent. Observations from 2002-2004 comprise the placebo cohorts, while observations from 2005-2007 comprise the reform cohorts.

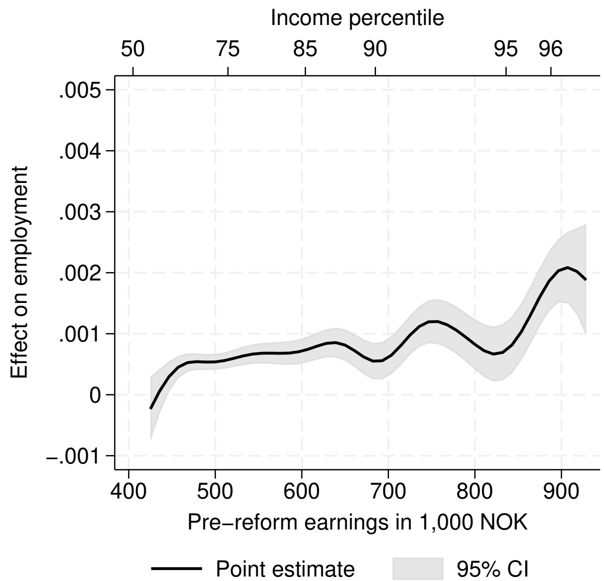


Figure 12: Local employment effects across initial income X.

Notes: The figure plots the local employment effects obtained by first estimating the counterfactual change in employment using 50 quantile bins of X, then estimating the numerator of equation (12) using a local regression, with the change in employment replacing the change in log incomes. 95 percent confidence intervals are shown, with standard errors obtained by bootstrapping the entire estimation procedure using 500 replications.

Table 5: Summary statistics of individual characteristics and labor market outcomes for lottery sample

	Winners	Population
Labor earnings	489,213	447,720
Employment	0.98	0.96
Age	45.01	42.20
Male	0.61	0.52
Married	0.55	0.55
Years of schooling	11.93	12.17
Household size	2.82	2.95
Q1 share	0.18	0.25
Q2 share	0.25	0.25
Q3 share	0.27	0.25
Q4 share	0.30	0.25

Notes: Monetary values are consumer-price-index adjusted and reported in 2018 NOK (8.13 NOK/USD). Winners' values are measured one year prior to the win and cohort-size weighted. Population statistics cover ages 25–61 from 1996–2017. Net worth is at the household level, normalized by the number of adults. The second panel reports the share of winners in each quartile of the annual earnings distribution of the working-age population; for the population, each share is mechanically 0.25.

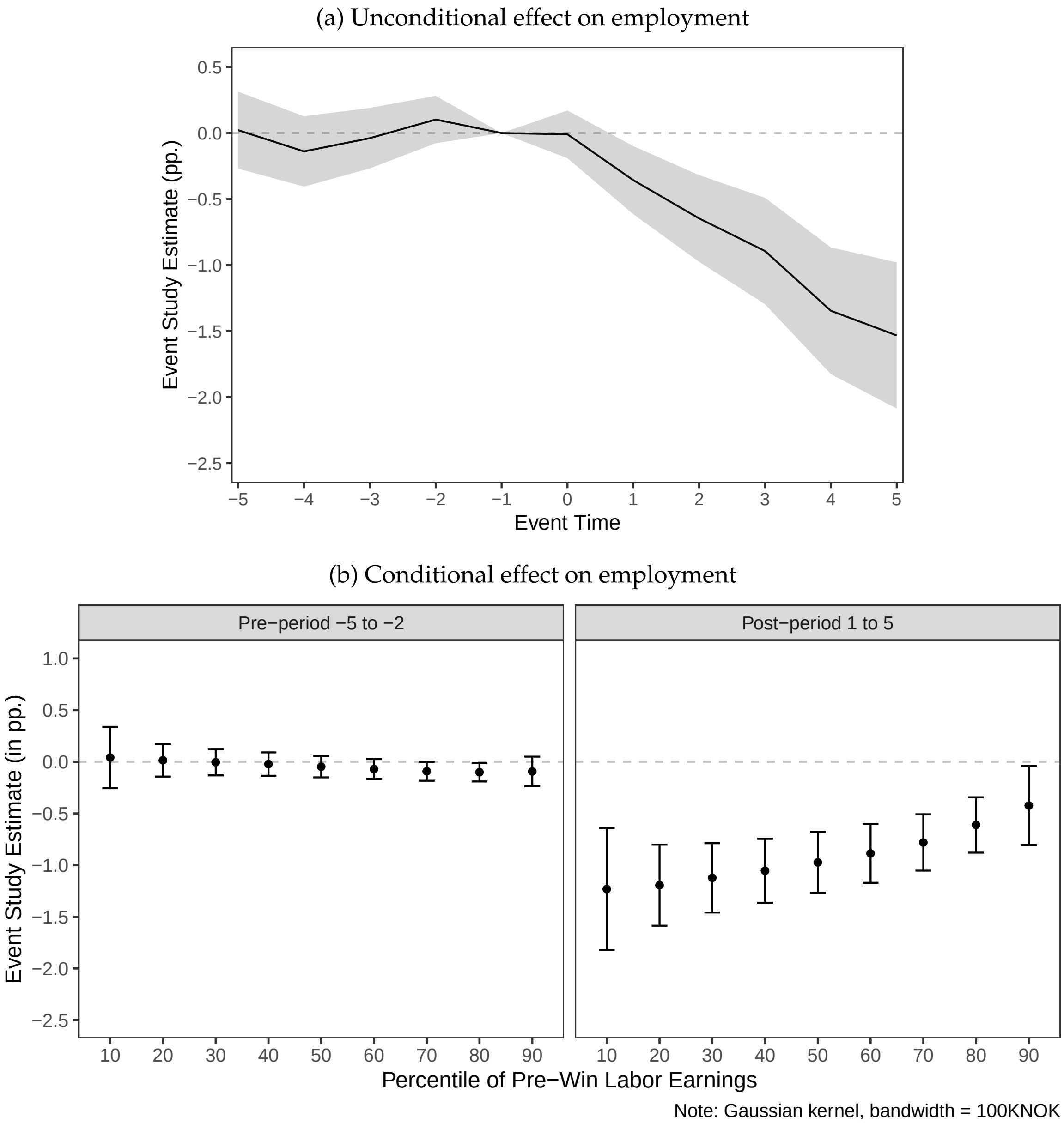


Figure 13: Employment effects of winning

Notes: This figure presents estimates of the effect of winning the lottery on winners' employment. Panel (a) presents unconditional estimates of the effect of winning the lottery on winners' employment for each event time t. The estimates correspond to the sample analogue of equation (25) using employment as an outcome instead of earnings, evaluated without conditioning on X, controlling for age, and averaged across cohorts using cohort-size weights. The estimates in Panel (b) correspond to the sample analogues of equation (25) using employment as an outcome instead of earnings, evaluated at different values of X. We use a Gaussian kernel with a bandwidth of 100,000 NOK, where the evaluation points $\{x\}$ correspond to the cohort-specific deciles of pre-win earnings. For each decile, we compute cohort-size weighted averages across cohorts and report equally weighted averages across event times $t = -5$ to -2 and $t = +1$ to $+5$. 95 percent confidence intervals are shown, with standard errors clustered at the winner level. Throughout, we use $g - 1$ as the omitted event time.

C Formal identification results for the lottery design

This appendix derives the formal identification results underlying our empirical strategy for the lottery design. We first map potential outcomes to observed data, show that the reduced-form estimand identifies the causal earnings effect under a conditional parallel-trends assumption, and then establish the decomposition of the total earnings response into intensive- and extensive-margin components.

Mapping potential outcomes to observed data. Let $W_{g,t}(d)$ denote the potential outcome W t years after winning ($d = 1$) or not winning ($d = 0$) the lottery in year $G = g$, for $W \in \{Y, B\}$, where Y and B denote earnings and unearned income, respectively. Potential outcomes map to the observed data through

$$W_{g+t} = \mathbb{1}[G \leq g+t]\, W_{G,t}(1) + \mathbb{1}[G > g+t]\, W_{g,t}(0). \tag{55}$$

Earnings, employment, and unearned income effects of winning the lottery. We start by stating the parallel trends assumption we rely on for obtaining the effect of winning the lottery on earnings and unearned income. It says that conditional on pre-win earnings $Y_{g-1} = x$, average earnings and unearned income for winners and later-winners would have evolved in parallel:

$$\begin{aligned} &\mathbb{E}[W_{g,t}(0) - W_{g,-1}(0) \mid Y_{g-1} = x,\, G = g] \\ &\qquad\qquad = \mathbb{E}\left[W_{g,t}(0) - W_{g,-1}(0) \mid Y_{g-1} = x,\, G > g+t\right], \end{aligned} \tag{56}$$

for all $t \geq 0$ and $W \in \{Y, B\}$. Under this assumption, the reduced form estimand $\text{RF}_{g,t}(x)$ ($\text{FS}_{g,t}(x)$) recovers the average earnings (unearned income) effect of winning the lottery for cohort g. To see why, note that

$$\begin{aligned} \text{RF}_{g,t}(x) &\equiv \mathbb{E}[Y_{g+t} - Y_{g-1} \mid Y_{g-1} = x, G = g] - \mathbb{E}[Y_{g+t} - Y_{g-1} \mid Y_{g-1} = x, G > g+t] \\ &= \mathbb{E}[Y_{G,t}(1) - Y_{g,-1}(0) \mid Y_{g-1} = x, G = g] - \mathbb{E}[Y_{g,t}(0) - Y_{g,-1}(0) \mid Y_{g-1} = x, G > g+t] \\ &= \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0) \mid Y_{g-1} = x, G = g] \\ &\quad + \Big(\mathbb{E}[Y_{g,t}(0) - Y_{g,-1}(0) \mid Y_{g-1} = x, G = g] - \mathbb{E}[Y_{g,t}(0) - Y_{g,-1}(0) \mid Y_{g-1} = x, G > g+t] \Big), \end{aligned}$$

for all $t \geq 0$. By the parallel-trends assumption (56), the term in parentheses equals zero, and hence

$$\text{RF}_{g,t}(x) = \mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0) \mid Y_{g-1} = x,\, G = g]\,.$$

Analogous arguments apply to unearned income and employment.

Aggregating across cohorts and event times. We aggregate the cohort and event-time specific reduced form and first stage estimands according to

$$\text{RF}(x) \equiv \sum_{g\in\mathcal{G}} \mathbb{P}(G=g \mid Y_{g-1}=x)\,\frac{1}{5}\sum_{t=1}^{5}\text{RF}_{g,t}(x), \tag{57}$$

$$\text{FS}(x) \equiv \sum_{g\in\mathcal{G}} \mathbb{P}(G=g \mid Y_{g-1}=x)\,\frac{1}{5}\sum_{t=1}^{5}\text{FS}_{g,t}(x). \tag{58}$$

The ratio between the two estimands then recovers

$$\frac{\text{RF}(x)}{\text{FS}(x)} = \sum_{g\in\mathcal{G}}\sum_{t=1}^{5}\omega_{g,t}(x)\,\frac{\mathbb{E}[Y_{g,t}(1)-Y_{g,t}(0) \mid Y_{g-1}=x, G=g]}{\mathbb{E}[B_{g,t}(1)-B_{g,t}(0) \mid Y_{g-1}=x, G=g]}, \tag{59}$$

with weights

$$\omega_{g,t}(x) = \frac{\mathbb{P}(G=g \mid Y_{g-1}=x)\;\mathbb{E}[B_{g,t}(1)-B_{g,t}(0) \mid Y_{g-1}=x, G=g]}{\sum_{j\in\mathcal{G}}\sum_{k=1}^{5}\mathbb{P}(G=j \mid Y_{g-1}=x)\;\mathbb{E}[B_{j,k}(1)-B_{j,k}(0) \mid Y_{g-1}=x, G=j]}. \tag{60}$$

The weights are positive and sum to one provided unearned income increases with the prize.

Recovering intensive-margin income effects To isolate the intensive-margin earnings response, we restrict attention to individuals who would be working t years after winning. We impose that conditional on pre-win earnings $Y_{g-1}=x$, individuals in cohort g who would be working at event time t would have experienced the same counterfactual earnings evolution as later winners who are observed to be working in year $g+t$:

$$\begin{aligned}&\mathbb{E}[Y_{g,t}(0)-Y_{g,-1}(0)\,|\,Y_{g-1}=x,\,G=g,\,Y_{g,t}(1)>0]\\ &\qquad\qquad = \mathbb{E}\left[Y_{g,t}(0)-Y_{g,-1}(0)\,|\,Y_{g-1}=x,\,G>g+t,\,Y_{g,t}(0)>0\right].\end{aligned} \tag{61}$$

Two features of our setting make this assumption plausible. First, conditioning on pre-win earnings absorbs most systematic differences in earnings capacity and labor-market attachment. Second, the extensive-margin response is small, so conditioning on being employed at $g+t$ selects nearly the same individuals in the treated and control groups. Together, these features substantially mitigate concerns about selection.

Under (61), the intensive-margin reduced form becomes

$$\begin{aligned}\text{RF}^{int}_{g,t}(x) &\equiv \mathbb{E}[Y_{g+t}-Y_{g-1} \mid Y_{g-1}=x, G=g, Y_{g+t}>0]\\ &\quad - \mathbb{E}[Y_{g+t}-Y_{g-1} \mid Y_{g-1}=x, G>g+t, Y_{g+t}>0]\\ &= \mathbb{E}[Y_{g,t}(1)-Y_{g,t}(0) \mid Y_{g-1}=x, G=g, Y_{g,t}(1)>0]\,.\end{aligned}$$

The corresponding intensive-margin first stage recovers

$$\text{FS}^{int}_{g,t}(x) = \mathbb{E}[B_{g,t}(1)-B_{g,t}(0) \mid Y_{g-1}=x, G=g, Y_{g,t}(1)>0]\,. \tag{62}$$

Our assumption of constant income effects conditional on Y_{g-1}, implies that

$$Y_{g,t}(1) = Y_{g,t}(0) + \frac{\eta(x)}{1-\tau} \left(B_{g,t}(1) - B_{g,t}(0)\right),$$

for individuals who work after winning. Substituting this into the intensive-margin reduced form $\text{RF}^{int}_{g,t}(x)$, we obtain

$$\frac{\text{RF}^{int}_{g,t}(x)}{\text{FS}^{int}_{g,t}(x)} = \frac{\eta(x)}{1-\tau}. \tag{63}$$

Decomposing the total earnings response. To decompose the total earnings effect per additional NOK of unearned income into the intensive- and extensive margin contributions, it is useful to introduce notation for being employed, $E_{g+t} = \mathbb{1}[Y_{g+t} > 0]$ and define the extensive-margin reduced form estimand by

$$\text{RF}^{ext}_{g,t}(x) \equiv \mathbb{E}[E_{g+t} - E_{g-1} \mid Y_{g-1} = x, G = g] - \mathbb{E}[E_{g+t} - E_{g-1} \mid Y_{g-1} = x, G > g+t]. \tag{64}$$

Under the parallel trends assumption and the additional assumption that employment decreases in lottery winnings, the extensive-margin reduced form estimand recovers:

$$\text{RF}^{ext}_{g,t}(x) = -\,\mathbb{P}(Y_{g,t}(0) > 0,\, Y_{g,t}(1) = 0 \mid Y_{g-1} = x, G = g). \tag{65}$$

By the law of total expectations, the total earnings effect can be written as:

$$\begin{aligned}
&\overbrace{\mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0) \mid Y_{g-1} = x, G = g]}^{\text{total effect}=\text{RF}_{g,t}(x)} \\
&\quad = \overbrace{\underbrace{\mathbb{P}(Y_{g,t}(1) > 0 \mid Y_{g-1} = x, G = g)}_{\text{employment share }=\mathbb{P}(Y_{g,t}>0|Y_{g-1}=x,G=g)} \times \underbrace{\mathbb{E}[Y_{g,t}(1) - Y_{g,t}(0) \mid Y_{g-1} = x, G = g, Y_{g,t}(1) > 0]}_{\text{intensive margin response}=\text{RF}^{int}_{g,t}(x)}}^{\text{intensive margin component}} \\
&\quad \underbrace{\overbrace{-\,\mathbb{P}(Y_{g,t}(1) = 0, Y_{g,t}(0) > 0 \mid Y_{g-1} = x, G = g)}^{\text{extensive margin response}=\text{RF}^{ext}_{g,t}(x)} \times \mathbb{E}[Y_{g,t}(0) \mid Y_{g-1} = x, G = g, Y_{g,t}(1) = 0, Y_{g,t}(0) > 0]}_{\text{extensive margin component}}.
\end{aligned}$$

Since there is only one unknown quantity in the expression, we can solve for it to obtain the average $Y_{g,t}(0)$ for those who would stop working if they won:

$$\mathbb{E}[Y_{g,t}(0) \mid Y_{g-1} = x, G = g, Y_{g,t}(1) = 0, Y_{g,t}(0) > 0] = \frac{\text{RF}_{g,t}(x) - p_{g,t}(x)\text{RF}^{int}_{g,t}(x)}{-\text{RF}^{ext}_{g,t}(x)} \tag{66}$$

where the employment share is denoted by $p_{g,t}(x) \equiv \mathbb{P}(Y_{g,t} > 0 \mid Y_{g-1} = x, G = g)$. Dividing the total effect by the first stage then yields:

$$\frac{\text{RF}_{g,t}(x)}{\text{FS}_{g,t}(x)} = \underbrace{\frac{p_{g,t}(x)}{\text{FS}_{g,t}(x)} \times \text{RF}^{int}_{g,t}(x)}_{\text{intensive-margin component}} + \frac{\overbrace{\text{RF}_{g,t}(x) - p_{g,t}(x)\text{RF}^{int}_{g,t}(x)}^{\text{extensive-margin component}}}{\text{FS}_{g,t}(x)}. \tag{67}$$

Since all the quantities in the expression are either identified or functions of the data, the decomposition is identified.

D Derivation of revenue-maximizing tax rate and marginal deadweight loss of income taxation with consumption and payroll taxes

D.1 Marginal deadweight loss

It is useful to start by noting that

$$\begin{aligned}
\frac{dDWL}{d\tau} &= Y - \sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^c}{\partial \tau} - Y - (\tau + \tau_w)\frac{\partial Y^c}{\partial \tau}, \\
&= -\left(\sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^c}{\partial \tau} + (\tau + \tau_w)\frac{\partial Y^c}{\partial \tau} \right),
\end{aligned}$$

where the first equality uses the envelope theorem. It is well-known from duality theory that the following equality between the compensated and uncompensated demand functions holds.

$$C_k^c(\tau_1, ..., \tau_K, \tau, \bar{V}) = C_k^u(\tau_1, ..., \tau_K, \tau, E(\tau_1, ..., \tau_K, \tau, \bar{V})),$$

where $E(.)$ is the expenditure function defined in equation (29). Two-step budgeting implies that,

$$C_k^c(\tau_1, ..., \tau_K, \tau, \bar{V}) = C_k(\tau_1, ..., \tau_K, I(\tau_1, ..., \tau_K, \tau, E(\tau_1, ..., \tau_K, \tau, \bar{V}))).$$

Taking the derivative of the compensated demand function with respect to τ thus yields,

$$\frac{\partial C_k^c}{\partial \tau} = \frac{\partial C_k^u}{\partial I}\left(\frac{\partial I}{\partial \tau} + \frac{\partial I}{\partial R}\frac{\partial E}{\partial \tau}\right).$$

By recalling that

$$I(\tau_1, ..., \tau_K, \tau, R) = (1-\tau)Y^u(\tau_1, ..., \tau_K, \tau, R) + R,$$

we obtain

$$\begin{aligned}
\frac{\partial I}{\partial \tau} + \frac{\partial I}{\partial R}\frac{\partial E}{\partial \tau} &= -Y + (1-\tau)\frac{\partial Y^u}{\partial \tau} + \left[(1-\tau)\frac{\partial Y^u}{\partial R} + 1\right]\frac{\partial E}{\partial \tau}, \\
&= -Y + (1-\tau)\frac{\partial Y^u}{\partial \tau} + \left[(1-\tau)\frac{\partial Y^u}{\partial R} + 1\right] Y, \\
&= -(1-\tau)\frac{\varepsilon^u Y}{1-\tau} + \eta Y, \\
&= -\varepsilon^c Y,
\end{aligned}$$

where the second-to-last equality uses the definition of the uncompensated elasticity and the last equality uses the Slutsky equation.

This implies that we can write the sum in the expression for the MDWL as

$$\sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^c}{\partial \tau} = -\varepsilon^c Y \sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k}{\partial I},$$
$$= -\varepsilon^c Y \frac{\tilde{\tau}}{1+\tilde{\tau}},$$

where the second equality uses the definition of $\tilde{\tau}$ from equation (28) and equation (68) below.

The marginal deadweight loss can now be written as

$$MDWL = \varepsilon^c Y \frac{\tilde{\tau}}{1+\tilde{\tau}} + (\tau + \tau_w)\frac{\varepsilon^c Y}{1-\tau},$$

which, after rearranging, yields the expression in the numerator of equation (31). To derive the marginal (compensated) tax revenue, we note that

$$\frac{\partial TR^c}{\partial \tau} = \sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^c}{\partial \tau} + Y + (\tau + \tau_w)\frac{\partial Y^c}{\partial \tau},$$
$$= Y - MDWL.$$

The expression in the denomiator of equation (31) follows immediately.

D.2 Revenue-maximizing tax rate

By standard arguments, increasing the top-income tax rate by $d\tau$ is equivalent to the marginal tax rate changing by $d\tau$ and the virtual transfer changing by $d\tau\bar{Y}$ in a linear tax system. The effect on government revenue can, therefore, be expressed as,

$$\frac{dTR}{d\tau_{\text{TOP}}} = \sum_{k=1}^{K} \tau_k P_k \left(\frac{\partial C_k^u}{\partial \tau} + \frac{\partial C_k^u}{\partial R}\bar{Y}\right) + (Y^u - \bar{Y}) + (\tau + \tau_w)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right).$$

Using the demand functions in equation (27), we obtain,

$$\frac{\partial C_k^u}{\partial \tau} + \frac{\partial C_k^u}{\partial R}\bar{Y} = \frac{\partial C_k^u}{\partial I}\left(-Y^u + \bar{Y} + (1-\tau)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right)\right).$$

Substituting this back into the expression above, we get that

$$\frac{dTR}{d\tau_{\text{TOP}}} = \left(-(Y^u - \bar{Y}) + (1-\tau)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right)\right)\sum_{k=1}^{K} \tau_k P_k \frac{\partial C_k^u}{\partial I}$$
$$+ (Y^u - \bar{Y}) + (\tau + \tau_w)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right).$$

Using the defintion of $\tilde{\tau}$ from equation (28), we have that,

$$1+\tilde{\tau} = \frac{\sum_{k=1}^{K}(1+\tau_k)P_k\frac{\partial C_k}{\partial I}}{\sum_{k=1}^{K}P_k\frac{\partial C_k^u}{\partial I}} = \frac{1}{\sum_{k=1}^{K}P_k\frac{\partial C_k^u}{\partial I}},$$

where the second equality follows since the budget constraint ensures that $\sum_{k=1}^{K}(1+\tau_k)P_k\partial C_k^u/\partial I = 1$. This means that

$$\sum_{k=1}^{K}P_k\frac{\partial C_k^u}{\partial I} = \frac{1}{1+\tilde{\tau}}.$$

which implies,

$$\sum_{k=1}^{K}\tau_k P_k\frac{\partial C_k^u}{\partial I} = \frac{\tilde{\tau}}{1+\tilde{\tau}}. \tag{68}$$

Using this, the revenue effect can be written as,

$$\begin{aligned}
\frac{dTR}{d\tau_{\text{TOP}}} &= \left(-(Y^u-\bar{Y}) + (1-\tau)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right)\right)\frac{\tilde{\tau}}{1+\tilde{\tau}} \\
&\quad + (Y^u-\bar{Y}) + (\tau+\tau_w)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right), \\
&= -(Y^u-\bar{Y})\frac{\tilde{\tau}}{1+\tilde{\tau}} + (Y^u-\bar{Y}) \\
&\quad + \left((\tau+\tau_w) + (1-\tau)\frac{\tilde{\tau}}{1+\tilde{\tau}}\right)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right), \\
&= \frac{Y^u-\bar{Y}}{1+\tilde{\tau}} + \left(\tau_w + \frac{\tau+\tilde{\tau}}{1+\tilde{\tau}}\right)\left(\frac{\partial Y^u}{\partial \tau} + \frac{\partial Y^u}{\partial R}\bar{Y}\right)
\end{aligned}$$

From the definitions of the uncompensated elasticity and income effect, we obtain that,

$$\frac{\partial Y^u}{\partial \tau} = -\frac{Y^u\varepsilon^u}{1-\tau}, \qquad \frac{\partial Y^u}{\partial R} = \frac{\eta}{1-\tau}.$$

Substituting this into the expression above,

$$\frac{dTR}{d\tau_{\text{TOP}}} = \frac{Y^u-\bar{Y}}{1+\tilde{\tau}} + \frac{\tau_w(1+\tilde{\tau})+\tau+\tilde{\tau}}{(1+\tilde{\tau})(1-\tau)}\left(-\varepsilon^u Y^u + \eta\bar{Y}\right)$$

At the revenue-maximizing tax rate, it must be the case that further increases in the top-income tax rate do not affect revenue. Thus,

$$\mathbb{E}\left[\frac{dTR}{d\tau_{\text{TOP}}} \mid Y \geq \bar{Y}\right] = \mathbb{E}\left[\frac{Y^u-\bar{Y}}{1+\tilde{\tau}} + \frac{\tau_w(1+\tilde{\tau})+\tau_{\text{TOP}}+\tilde{\tau}}{(1+\tilde{\tau})(1-\tau_{\text{TOP}})}\left(-\varepsilon^u Y^u + \eta\bar{Y}\right) \mid Y \geq \bar{Y}\right] = 0,$$

where the expectation is taken across individuals in the top-income tax bracket. Using the definitions from equation (34) and the assumption that $\tilde{\tau}$ is constant across individuals, we obtain,

$$\begin{aligned}\frac{\mathbb{E}[Y^u - \bar{Y} \mid Y^u \geq \bar{Y}]}{1+\tilde{\tau}} + \frac{\tau_w(1+\tilde{\tau}) + \tau_{\text{TOP}} + \tilde{\tau}}{(1+\tilde{\tau})(1-\tau_{\text{TOP}})}\left(-\bar{\varepsilon}^u\, \mathbb{E}[Y^u \mid Y^u \geq \bar{Y}] + \bar{\eta}\bar{y}\right) &= 0,\\ \frac{(\alpha-1)\bar{y}}{1+\tilde{\tau}} + \frac{\tau_w(1+\tilde{\tau}) + \tau_{\text{TOP}} + \tilde{\tau}}{(1+\tilde{\tau})(1-\tau_{\text{TOP}})}\left(-\bar{\varepsilon}^u \alpha\bar{y} + \bar{\eta}\bar{y}\right) &= 0,\\ (\alpha-1) + \frac{\tau_w(1+\tilde{\tau}) + \tau_{\text{TOP}} + \tilde{\tau}}{1-\tau_{\text{TOP}}}\left(-\bar{\varepsilon}^u \alpha + \bar{\eta}\right) &= 0\end{aligned}$$

where the second line uses the defintion of α and the third multiplies both sides by $(1+\tilde{\tau})/\bar{Y}$. Rearranging, we obtain,

$$\frac{\tau_w(1+\tilde{\tau}) + \tau_{\text{TOP}} + \tilde{\tau}}{1-\tau_{\text{TOP}}} = \frac{\alpha-1}{\bar{\varepsilon}^u\alpha - \bar{\eta}}.$$

Solving for τ_{TOP}, we arrive at

$$\tau_{\text{TOP}} = \frac{\alpha - 1 - (\tilde{\tau} + \tau_w + \tau_w\tilde{\tau})\,(\alpha\bar{\varepsilon}^u - \bar{\eta})}{\alpha - 1 + (\alpha\bar{\varepsilon}^u - \bar{\eta})}.$$

E Dynamic labor supply model

This appendix shows how the static model in Section 4 relates to a standard dynamic life-cycle model following Blomquist (1985), Blundell and Walker (1986), and MaCurdy (1983). For simplicity, we abstract from non-linear labor income taxation, assume that the return on capital is fixed over time and focus on the case where there is only one consumption good and one margin of labor supply per period.

Consider an individual who won ξ dollars at time g with wealth A_{g-1}. They solve:

$$\max_{\{C_t,Y_t,A_t\}_{t=g}^{T}} \sum_{t=g}^{T} \beta^t U(C_t, Y_t), \quad \text{subject to} \quad C_g + A_g = (1-\tau)Y_g + (1+r)A_{g-1} + \xi,$$
$$C_t + A_t = (1-\tau)Y_t + (1+r)A_{t-1}, \quad \text{and} \quad A_T \geq 0 \quad \text{for all } t > g. \tag{69}$$

The solution to the problem depends on the lottery winnings, so we write $C_t(\xi), Y_t(\xi)$ and $A_t(\xi)$ for all $t \geq g$. The following first-order condition determines labor supply in period t:

$$\frac{\partial U}{\partial C}(1-\tau) + \frac{\partial U}{\partial Y} = 0,$$

where the derivatives are evaluated for consumption $C = (1-\tau)Y_t(\xi) + B_t(\xi)$ with $B_t(\xi) \equiv (1 + r)A_{t-1}(\xi) - A_t(\xi) + \mathbb{1}[t = g]\xi$ and earnings $Y = Y_t(\xi)$. This means we can write labor supply as

$$Y_t(\xi) = Y(\tau, B_t(\xi)),$$

The effect at time t of winning an additional dollar is given by

$$\frac{dY_t}{d\xi} \quad \text{and} \quad \frac{dB_t}{d\xi}.$$

Moreover, by the chain rule

$$\frac{dY_t}{d\xi} = \frac{\partial Y}{\partial B}\frac{dB_t}{d\xi} = \frac{\eta}{1-\tau}\frac{dB_t}{d\xi},$$

where η is the same parameter as in the static model in Section 4. Thus, the dynamic model implies the income effect is related to the earnings and unearned income effect of lottery winnings through,

$$\frac{\partial Y_t/\partial \xi}{\partial B_t/\partial \xi} = \frac{\eta}{1-\tau}.$$