The Norwegian Labor Income Tax Credit Experiment – Research projects

October 6, 2026

1 Introduction

The Norwegian Labor Income Tax Credit Experiment began in early 2026. Bensnes et al. (2026) documents the implementation of the experiment and describes the planned reporting of results to the Ministry of Finance. In parallel, the research group behind the report, together with researchers from other institutions, is undertaking several research projects that build on the experiment. This document describes these projects.

2 Background, experimental design and data

Background. The Norwegian Labor Income Tax Credit Experiment is a large randomized experiment embedded in the national tax system. The experiment was developed jointly by the Ministry of Finance, the Norwegian Tax Administration, and researchers at Norwegian Fiscal Studies at the University of Oslo, the University of Chicago, the University of Stavanger, and the Frisch Centre. The Norwegian Tax Administration is responsible for its implementation.

The experiment addresses a long-standing policy question: how do changes to the tax system affect labor supply? A labor-income tax credit has been proposed several times as a way to strengthen work incentives among people otuside the labor force or with low earnings. It was proposed by the Solberg government in 2021, and recommended by a public commission in 2022 (NOU 2022:20). Such a credit can make it more attractive to enter employment or increase earnings. Its' overall effect on labor supply is nevertheless ambiguous. Higher after-tax income may lead some people who already work to reduce their labor supply. The phase-out also lowers the return to additional earnings. The size and balance of these responses are empirical questions.

The tax credit. The credit applies from the 2026 income year and is planned to remain in place through 2030. The calculation base consists mainly of labor earnings and self-employment income. It excludes work assessment allowance, disability benefits, and

unemployment benefits. Some benefits that replace labor earnings, including sickness and parental benefits, are included.

In 2026, the deduction is at most NOK 125,000 and can reduce annual tax liability by up to NOK 27,500. Table 1 summarizes how the credit changes tax rates for an individual with only labor earnings and standard deductions. For such an individual, the credit is phased in at earnings of about NOK 210,000 and is fully used at about NOK 335,000. The deduction is phased out at a rate of 40 percent once earnings exceed NOK 345,000 and is fully phased out at NOK 657,500. This means that above NOK 657,500, the experimental and baseline tax schedules coincide. The nominal parameters are expected to be adjusted annually in line with wage growth.

Table 1: Incentive regions of the tax credit

Region	Earnings range (NOK)	Marginal tax rate	Average tax rate	Participation tax rate
Phase-in	210,000–335,000	↓	↓	↓
Plateau	335,000–345,000	—	↓	↓
Phase-out	345,000–657,500	↑	↓	↓
Above	Above 657,500	—	—	—

Table 1 also illustrates how the tax credit changes work incentives. It shows that the credit can affect labor supply along two margins: the extensive margin—whether to work—and the intensive margin—how much to work conditional on working. The tax credit decreases participation tax rates, defined here as the share of earnings lost to taxes upon entering employment, for earnings between NOK 210,000 and NOK 657,500 and therefore strengthens incentives to enter employment. In the phase-in region, the tax credit decreases both marginal and average tax rates, so income and substitution effects work in opposite directions when leisure is a normal good. In the phase-out region, it continues to reduce average taxes but raises the marginal tax rate, suggesting that it weakens the incentives to work on the intensive margin.

Experimental design. The experiment started in January 2026 and is planned to run until the end of 2030. The target population consists of approximately 1.33 million individuals born in 1991–2006 who were randomized into three groups:

- **Group A (8 percent)** faces the experimental labor-income tax credit and receives personalized information about work incentives under the experimental schedule.
- **Group B (46 percent)** remains under the baseline schedule and receives personalized information about work incentives under that schedule.
- **Group C (46 percent)** remains under the baseline schedule and receives no personalized incentive information.

For the tax credit to affect behavior, individuals must know whether it applies to them. We now explain how the experimental design allows us to estimate its effect among individuals who were informed of their assignment.

When the 2026 tax deduction cards became available in December 2025, members of groups A and B received the following notification by text message or email:

> "X's tax withholding card for 2026 is ready. This year, the Government is launching a trial scheme regarding an employment deduction for young people. Log in to our website to check if you are eligible for this deduction. Remember to check that the information in your tax deduction card is correct."

After logging in to the Norwegian Tax Administration's website, both groups saw the same banner linking to the experiment's information page. No assignment-specific information was shown before they opened the page. After clicking the link in the banner, visitors were told whether they had been assigned to the tax-credit schedule. They were also shown key figures from their tax deduction card and how higher labor earnings would affect taxes and after-tax income. The examples considered earnings increases of NOK 1,000, NOK 100,000, and NOK 300,000. For most individuals, the calculations were tailored using information from their tax deduction card. Group A saw calculations under the tax-credit schedule, while group B saw calculations under the baseline tax rules. Individuals receiving social-security benefits instead saw two standardized examples.[1]

Everyone who opened the page was therefore informed of their assignment. Because groups A and B had the same experience until the page was opened, the decision to open it could not depend on assignment. Random assignment is therefore preserved among page visitors. We refer to these individuals in groups A and B as the *informed sample*. Throughout the projects, experimental comparisons of the tax schedules use the informed sample, comparing group A with group B. Survey analyses use respondents within the informed sample.

Data sources. All projects will use several registry databases maintained by Statistics Norway and link them with the assignment records. This provides us with a population panel data set with detailed information on Norwegian individuals, families, and firms over a long period of time. The registry databases contain records for every individual and family in Norway from 1967 onward. The register data enumerate individual demographic information (including sex, age, immigrant status, and education) and all sources of annual income, including earnings, self-employment income, capital income, and cash transfers. Complete records of individuals' participation in social support programs such as unemployment benefits, disability insurance, and social assistance are available from 1992 onward.

The project will also rely on consumption data from two sources. The first uses

[1] Personalized calculations for benefit recipients would require information held by the Norwegian Labour and Welfare Administration but not by the Norwegian Tax Administration.

data on income, capital gains, and changes in wealth to impute household consumption expenditure following Eika et al. (2020). The second uses high-frequency, person-level payment data from Nets Branch Norway, which provide information on the timing and product categories of purchases

These data will be linked to survey measures of tax and benefit knowledge, beliefs about returns to work, desired labor supply, work–study constraints, institutional trust, tax morale, and fairness.

3 Labor-supply responses to a randomized income-tax credit

Policymakers and researchers are keenly interested in how to design transfer systems for low-income individuals. Some advocate tax credits designed to increase labor-force participation at the cost of weakening incentives to earn more. Others favor a universal basic income or negative income tax, which provides broader income support but may weaken participation incentives. The costs and benefits of these alternatives depend on labor-supply responses along the extensive and intensive margins and on the relative importance of income and substitution effects. As discussed in Section 3.2, there is no consensus on the magnitude of these responses, making it difficult to design evidence-based tax and transfer policies for low-income individuals. This motivates our project, which uses the tax experiment to answer the following main questions:

Q1: What is the mean impact of the tax credit?

Existing evidence on the labor-supply impacts of tax credits relies on quasi-experimental methods that require untestable assumptions. The tax credit experiment allows us to provide credible evidence on how average employment, earnings, hours worked, and tax revenue are affected by the tax credit. It is, to the best of our knowledge, the first experiment that randomizes access to an earned income tax credit without simultaneously introducing additional policies that may also affect labor supply. This means that the estimated responses to the tax credit reflect only pure financial incentives and are not conflated with work requirements and time limits.

Another advantage of the setting is that the experimental population includes all individuals born in 1991–2006. The broad population lets us study labor-supply responses for both men and women, couples and singles, and households with and without children.

Q2: What is the distributional impact of the tax credit?

As pointed out by Bitler et al. (2006), mean impacts of the credit may miss a great deal: standard economic theory predicts rich heterogeneity in responses because the tax credit affects incentives differently throughout the income distribution. For example, individuals in the phase-out region experience increases in marginal and decreases in average tax rates, while those in the phase-in region experience decreases in both. Individuals who do

not work under the baseline may start working because of the tax credit. Mean impacts will average together positive and negative labor supply responses, possibly obscuring the extent of the tax credit's effects.

To begin to understand this heterogeneity, we will estimate how the tax credit changes the distribution of earnings and other outcomes by computing quantile treatment effects.

Our sample size allows us to estimate the distributions with and without the tax credit very precisely. More than 100,000 individuals receive the tax credit, while more than 600,000 receive the same information but no tax credit. By contrast, the quantile treatment effects in Bitler et al. (2006) are estimated using a total sample of fewer than 5,000 individuals. The use of administrative data also ensures that outcomes are measured with minimal error.

Q3: How important are the intensive and extensive margins, and how elastic is labor supply to changes in work incentives?

While it is interesting to know the quantile treatment effects of the tax credit, their implications for the design of the tax and transfer system are not immediate. What we really need to know is the distribution of effects, not the effects on the distribution. The efficiency and revenue consequences of different tax policies depends on knowledge about the distribution of the effects, especially the labor supply elasticities, both on the intensive and extensive margins.

To learn about these elasticities, we consider a rich nonparametric labor-supply model that allows for unrestricted heterogeneity in preferences, income effects, and fixed costs of working. Even though the model is flexible, it imposes restrictions on choices under the two tax systems and therefore on the possible distribution of treatment effects. We show how to combine these restrictions with data to construct bounds and point estimates of earnings responses on the intensive and extensive margins, as well as substitution and income effects.

Q4: What are the welfare and revenue consequences of counterfactual tax credits?

The experiment provides direct evidence on the labor-supply and revenue consequences of the implemented credit, but not on outcomes under counterfactual tax credits. To study the impact of such counterfactual tax credits, we combine the experimental evidence with the model developed for answering Q3. We examine how changing the phase-in and phase-out rates, as well as the size of the credit, affects intensive- and extensive-margin labor supply, revenue, and welfare. This provides evidence relevant to the design of the tax credit.

3.1 Methodology and model

We now explain the methodology and model we will use to answer questions 1–4 in more detail.

Experimental effects

The answers to questions 1 and 2 depend on the marginal distributions of earnings with and without the tax credit. Random assignment implies these distributions are identified from the data: for each outcome, the distribution in the group assigned the credit identifies its distribution under the experimental schedule, while the distribution in the control group identifies its distribution under the baseline schedule. We use these distributions to estimate effects on employment, earnings, hours, taxes, transfers, after-tax income, and other outcomes specified in the pre-analysis plan.[2] For each outcome Y and period, we estimate the following regression comparing groups A and B:

$$Y_i = \beta D_i + \sigma_{s(i)} + u_i,$$

where D_i denotes assignment to the tax credit and $\sigma_{s(i)}$ is a fixed effect for individual i's randomization stratum. The coefficient β identifies the mean effect of the credit.

We examine heterogeneity by predetermined covariates—including sex, age, student status, benefit receipt, and prior earnings—and by predicted exposure to the incentive regions in Table 1.

To answer Question 2, we compute quantile treatment effects, which describe how the tax credit changes different parts of each outcome's marginal distribution and can reveal effects obscured by comparisons of means (Bitler et al., 2006). We follow Firpo (2007) in how we estimate these quantile treatment effects.

While quantile treatment effects may reveal interesting patterns of response heterogeneity, they cannot be interpreted as a distribution of individual treatment effects (Heckman et al., 1997). The reason is that the two marginal distributions are consistent with many different joint distributions. Without further assumptions, we therefore cannot identify, for example, the earnings response among people who would work under either schedule, the shares induced to enter or exit employment, or the share who shift earnings into a region affected by the credit. These questions depend on the joint distribution of the two potential outcomes. Answering questions 3 and 4 therefore requires restrictions on this joint distribution.

Labor supply elasticities and counterfactual policy

A nonparametric labor-supply model. To impose these restrictions and provide the basis for counterfactual analysis, we consider a nonparametric labor-supply model. Under

[2] See https://www.socialscienceregistry.org/trials/17105.

tax schedule $d \in \{0, 1\}$, an individual chooses consumption C and earnings Y by solving

$$\max_{C,Y} \; u(C, Y) - \xi \mathbf{1}\{Y > 0\} \qquad \text{subject to} \qquad C = Y - T_d(Y),$$

where $u(C, Y)$ is increasing in consumption and decreasing in earnings, $\xi \geq 0$ is a fixed utility cost of working, and T_0 and T_1 denote the baseline and experimental tax schedules. The resulting choices Y_0 and Y_1 are earnings without and with the tax credit.

The nonparametric utility function allows for income effects, while the fixed cost permits discontinuous changes in labor-force participation. Both may vary arbitrarily across individuals. This unrestricted heterogeneity allows us to accommodate responses driven by consumption–leisure preferences, information frictions, adjustment costs, or a combination of these mechanisms (Chetty, 2012; H. J. Kleven & Waseem, 2013; Kostøl & Myhre, 2021).

Distinguishing these mechanisms matters for policy and for interpreting labor-supply responses, as discussed in Section 4. Our identification results, however, do not require us to separate them. In this project, we therefore study responses to actual tax changes, which may affect behavior through changes in perceived incentives.

Identification. The model serves two purposes. First, it defines the income and substitution effects on the intensive and extensive margins in Q3 and links them to the welfare and revenue consequences in Q4. Second, it restricts the joint distribution of Y_0 and Y_1 by ruling out pairs of earnings choices that cannot be generated by optimizing behavior. These support restrictions may help tighten bounds on target parameters that depend on the joint distribution.

To see how the labor-supply model imposes restrictions on Y_0 and Y_1, it is useful to consider a few examples. First, recall that the tax credit weakly increases the consumption available at any positive earnings level while leaving consumption at zero earnings unchanged. A person who prefers working under the baseline schedule therefore continues to prefer working under the credit. Furthermore, because the two schedules coincide above NOK 657,500, earnings chosen above this threshold under the credit remain optimal under the baseline schedule. Additional assumptions on preferences rule out further pairs of choices: Normality of consumption and leisure restricts income effects, while quasilinear utility eliminates income effects altogether.

These examples illustrate some model-implied restrictions on choices. However, they do not exhaust all of the model's implications. To provide a reliable way of extracting all the restrictions from the model, we will use the methods developed by Dutz and Håvarstein (2026), which build on the revealed-preference logic of Afriat (1967) and its extension to general and nonlinear budget sets (Cherchye et al., 2014; Forges & Minelli, 2009). These methods work by deriving necessary and sufficient conditions for any candidate earnings pair (y_0, y_1) to be consistent with utility maximization under a given set of assumptions on preferences. Dutz and Håvarstein (2026) show that these conditions

can be written as a set of linear inequalities that extract all the restrictions that the model places on the joint distribution of earnings.

To identify the labor supply parameters in Q3, we combine these restrictions with the marginal distributions identified by the experiment. A candidate joint distribution of (Y_0, Y_1) must be consistent with the model and reproduce the marginal distributions identified by the experiment. A large class of target parameters can be expressed as linear or linear-fractional functions of this joint distribution, meaning that their minimum and maximum values can be computed using linear programming (Russell, 2021; Torgovitsky, 2019). Because the revealed-preference conditions are necessary and sufficient, the feasible set coincides with the identified set, and the resulting bounds are sharp.

Counterfactual tax credits. To answer Q4, we use the intensive- and extensive-margin elasticities estimated in Q3 to assess how alternative phase-in rates, phase-out rates, and credit amounts affect labor supply and tax revenue. We combine these responses with welfare weights to evaluate the welfare consequences of each policy. When the elasticities are only partially identified, we report corresponding bounds on the counterfactual effects. We examine how the conclusions change with the economic assumptions imposed and when adjustment frictions are allowed.

3.2 Relation and contribution to existing work

Empirical work. There is little consensus on how transfers to low-income individuals affect labor supply and tax revenue. For example, Heckman (1993) argues that existing evidence suggests that labor supply is relatively elastic on the extensive margin: changes in incentives induce people to start or stop working. This view is consistent with experimental evidence from the Negative Income Tax and Jobs First experiments, which found substantial participation responses to changes in work incentives (Bitler et al., 2006; Kline & Tartari, 2016; P. K. Robins, 1985). It is also supported by later quasi-experimental evidence from expansions of the Earned Income Tax Credit in the US (Eissa & Hoynes, 2004; Eissa & Liebman, 1996; Meyer & Rosenbaum, 2001).

However, subsequent work has raised concerns about selective attrition and the small sample sizes in the Negative Income Tax experiments (Angrist & Pischke, 2010; Hausman & Wise, 1979; P. K. Robins & West, 1986). More recently, Riddell and Riddell (2024) document failures of random assignment in two experiments and show that the conclusions differ across experiments and by baseline welfare status. Furthermore, H. Kleven (2024) points out that most randomized evaluations of welfare programs bundle lower tax rates with time limits, work requirements, and other non-pecuniary work incentives. This makes it difficult to infer how much of the response is due to pure financial incentives. H. Kleven (2024) also argues that estimated responses to EITC expansions are fragile to specification choices and that the effects of the 1993 expansion align more closely with other reforms and macroeconomic shocks than with the EITC.

The experience with the introduction of an earned income tax credit in Sweden illustrates a related problem. Because all working-age individuals were covered by the earned income tax credit introduced in 2007, there was no untreated comparison group. Edmark et al. (2016) attempt to exploit differences in credit size across municipalities and predicted earnings, but conclude that "the question of whether the reform has delivered the hoped-for effects cannot be credibly answered."

This project contributes to this literature by providing credible evidence from a randomized tax credit experiment that does not change non-pecuniary incentives for work. We will provide novel estimates of intensive- and extensive-margin responses without relying on parametric assumptions about unobservables. The project will also deliver estimates of intensive- and extensive-margin substitution and income effects while allowing for unrestricted heterogeneity across individuals. These elasticities can, in turn, be combined with results from a large theoretical literature on optimal taxation to guide the design of the transfer system for low-income individuals (Bierbrauer et al., 2023; Jacquet et al., 2013; Laroque, 2005; Saez, 2001, 2002).

Methodological work. The methods used to answer Q3 and Q4 build on three strands of revealed-preference research. The first is the literature that uses revealed preference arguments to bound consumer demand: In their empirical application, Blundell et al. (2008) use repeated cross-sections and observational price variation to estimate semiparametric average Engel curves and bound demand under a new linear budget. The average demand responses may vary with expenditure and observed household characteristics, but the analysis does not recover a distribution of individual preferences. It also assumes exogenous prices, stable preferences, unique demand, and weak normality. Blundell et al. (2015) build on this approach and use transitivity to obtain sharp bounds on demand and welfare.

We instead use randomized variation in nonlinear and nonconvex tax schedules and model unrestricted individual heterogeneity in preferences, income effects, and fixed costs of working. We also compare the identifying content of several economic assumptions rather than maintaining one preference class throughout.

The second strand of literature considers partial identification of labor-supply models. Manski (2014) studies general nonlinear tax schedules and allows individuals to have heterogeneous nonparametric preferences and discrete or continuous labor-supply opportunities. Because each person is observed under only one schedule, population inference across choice sets requires assumptions about the comparability of preference distributions across groups. Manski explicitly points to randomized tax schedules as a way to strengthen identification. Kline and Tartari (2016) provide the closest experimental precedent: they combine a randomized welfare reform with a nonparametric optimizing model that allows fixed costs of working, and use linear programming to bound the probabilities of nine qualitative response types and aggregates of these types.

Our experiment identifies the marginal earnings distribution under each tax schedule

by design. We combine this variation with necessary and sufficient conditions for nonlinear and nonconvex tax schedules, heterogeneous fixed costs, and, in an extension, adjustment frictions. Compared to Kline and Tartari (2016), our approach allows us to consider identification of a much larger class of target parameters, including earnings responses and substitution and income effects. It also ensures that the resulting bounds are sharp, in the sense that they cannot be tightened without imposing additional assumptions or using more data.

Finally, the project also relates to the literature that uses revealed preference arguments to test economic models. Hoderlein and Stoye (2014) and Kitamura and Stoye (2018) ask whether observed choice distributions can be rationalized by utility-maximizing behavior.

We use the same revealed-preference logic for a different purpose. The experiment identifies the marginal earnings distributions under the two schedules but not their joint distribution. We characterize which couplings of these marginals are rationalizable and optimize each target parameter over the admissible couplings. The model remains testable: if no admissible coupling exists, the data reject the maintained economic assumptions. We conduct this specification check before drawing structural conclusions.

4 Information frictions in taxation and labor supply

The results from Section 3 do not necessarily tell us why individuals respond. Observed responses can reflect preferences for consumption and leisure or adjustment costs, but they can also reflect imperfect information about the actual incentives. In fact, a growing body of evidence suggests there are systematic errors in how workers perceive average and marginal tax rates (Gideon, 2017; Rees-Jones & Taubinsky, 2020). If perceived incentives differ from actual incentives, observed responses to tax reforms therefore combine two mechanisms: how reforms change perceptions and how labor supply responds to those perceptions. The goal of this project is to disentangle these two mechanisms. This is important because it tells us whether observed responses reflect underlying labor-supply elasticities or imperfect knowledge of tax incentives.

In this project, we combine the tax-credit experiment with a large survey experiment that randomly assigns individuals to receive personalized information about their marginal tax rate. We elicit their beliefs and desired labor supply and link the survey to administrative records on hours and earnings. We use these data to answer the following four questions:

Q1: Do workers understand their work incentives?

Existing evidence documents systematic errors in perceived marginal tax rates and suggests that many taxpayers approximate nonlinear tax schedules using average rather than marginal rates (Gideon, 2017; Rees-Jones & Taubinsky, 2020). We elicit each respondent's subjective distribution of marginal tax rates and compare it with the incentives implied by

the tax and transfer system. The information experiment then identifies how personalized information changes these beliefs and how updating depends on prior uncertainty. This allows us to study beliefs, uncertainty, and updating for the same individuals.

For biased beliefs to matter for actual outcomes, they have to affect behavior. Workers may update their beliefs about incentives without changing their labor supply. This motivates our second question.

Q2: How do perceived work incentives affect labor supply?

Existing evidence on whether tax information changes labor supply is mixed (Chetty & Saez, 2013; Chetty et al., 2013; Dur et al., 2026; Kostøl & Myhre, 2021). Our setting is uniquely well-suited to answer this question because it combines two experiments. The information experiment generates variation in perceptions while holding the tax schedule fixed, whereas the tax-credit experiment changes actual incentives and may also change perceptions. We use both experiments as instrumental variables when estimating how perceived incentives affect desired labor supply and observed hours and earnings.

The IV estimates of the effect of perceived incentives on labor supply do not reveal who remains uninformed or whether information frictions are concentrated among workers who are particularly responsive to incentives. This motivates our next question.

Q3: Who faces the largest information frictions, and how responsive are they to work incentives?

The information intervention and the tax-credit experiment may affect the beliefs of different workers differently. In the model we consider, the information intervention lowers the cost of becoming informed, while the tax credit changes actual incentives and the value of information. We use this model to interpret the IV estimates of the effect of perceived marginal tax rates on labor supply obtained using the two experiments as separate instruments. This allows us to characterize who faces the largest information frictions and for whom changing incentives and correcting mistaken beliefs is most likely to change behavior.

Q4: How do information frictions affect the distribution of earnings?

Labor supply choices depend on perceived incentives, and the labor supply elasticities we estimate when answering Q2 are with respect to perceived marginal tax rates. We use these elasticities to compute the earnings distribution that would emerge if perceived tax rates are set to equal the true marginal tax rate. This allows us to quantify the whole distributional effect of information frictions on labor supply and tax revenue.

Model and methodology

We now explain in more detail the methodology and model we will use to answer questions 1–4.

A labor-supply model with imperfect information. We organize the empirical analysis around a model in which workers hold subjective initial beliefs about their tax incentives and choose their earnings and whether to acquire information about the incentives they face. We focus on workers with strong attachments to the labor market, allowing us to ignore fixed costs of working. Let D denote assignment in the tax-credit experiment, let Z denote assignment in the survey information experiment, and let F_d denote an individual's beliefs about the marginal tax rate τ and virtual income R under $D = d$. Given their initial beliefs, the individual chooses earnings Y to maximize expected utility,

$$\max_{Y \geq 0} \mathbb{E}_{F_d}\left[(1-\tau)Y + R - \frac{\alpha}{1+1/\rho}\left(\frac{Y}{\alpha}\right)^{1+1/\rho}\right],$$

where the expectation is taken with respect to their subjective distribution F_d. The underlying quasi-linear utility function is used extensively in the public finance literature; see, e.g., Saez et al. (2012).

Optimal earnings depend on d only through the expected marginal tax rate $\bar{\tau}(d) \equiv \mathbb{E}_{F_d}[\tau]$:

$$Y(\bar{\tau}(d)) = \alpha\,(1-\bar{\tau}(d))^{\rho},$$

where ρ is the elasticity of earnings with respect to the perceived net-of-tax rate. The model allows beliefs, elasticities, and earnings capacity to vary across individuals.

Individuals can choose to become fully informed by incurring a cost $\xi_{\text{INFO}}(z)$ of attending to and processing information. The cost depends on assignment $Z = z$ in the information experiment. The value of full information is

$$V(F_d) = \frac{\alpha}{1+\rho}\left[\mathbb{E}_{F_d}\left[(1-\tau)^{1+\rho}\right] - (1-\bar{\tau}(d))^{1+\rho}\right].$$

The individual therefore chooses to acquire information when $V(F_d) \geq \xi_{\text{INFO}}(z)$. The perceived marginal tax rate after this decision is

$$\tau(D,Z) = \mathbf{1}[V(F_D) \geq \xi_{\text{INFO}}(Z)]\tau^*(D) + (1 - \mathbf{1}[V(F_D) \geq \xi_{\text{INFO}}(Z)])\,\bar{\tau}(D),$$

where $\tau^*(d)$ is the individual's true marginal tax rate under assignment $D = d$. This links information acquisition to belief uncertainty, labor-supply responsiveness, and the cost of processing information.

Beliefs and information updating (Q1). We compare respondents' subjective distributions of marginal tax rates with the rates implied by the tax and transfer system. This

measures both belief errors and subjective uncertainty. Random assignment in the information experiment identifies whether personalized information corrects these errors and whether updating depends on prior uncertainty. We also examine how beliefs and updating vary across individuals.

Perceived incentives and labor supply (Q2). We first use random assignment in the information experiment to estimate effects on beliefs, desired labor supply, and subsequent hours and earnings. We also consider how the tax credit affects the same outcomes. By using the two experiments as instruments for perceived incentives, we estimate labor-supply elasticities with respect to perceived marginal tax rates. As a validation exercise, we compare stated and realized responses to the information treatment, compare stated responses to actual and hypothetical incentive changes, and test whether preferences recovered from stated choices predict the experimental earnings distributions under both tax schedules.

Information frictions and heterogeneous responses (Q3). The two experiments may shift the beliefs of different workers. In the model, assignment in the information experiment changes the cost of processing information, whereas assignment in the tax-credit experiment changes actual incentives, beliefs, and the value of becoming informed. Under appropriate exclusion and directional-monotonicity conditions, each experiment therefore identifies a differently weighted average response to perceived incentives (Angrist & Imbens, 1995). Comparing these responses helps us learn how information costs, prior uncertainty, and labor-supply responsiveness vary together. We examine how the implied frictions vary with observed characteristics and identify the workers for whom correcting beliefs is most likely to affect behavior.

Effect of information frictions on earnings and tax revenue (Q4). To understand how we identify the earnings distribution that would emerge if perceived tax rates are set to equal the true marginal tax rate, it is useful to express individual elasticities as

$$\rho_i = h_\rho(X_i) + \epsilon_i,$$

where X_i is a vector of observable characteristics, $h_\rho(x) = E[\rho_i \mid X_i = x]$, and the residual ϵ_i satisfies $E[\epsilon_i \mid X_i] = 0$. This decomposition is without loss of generality. For simplicity, we rule out unobserved heterogeneity by setting $\epsilon_i = 0$ for each individual. We will extend this argument to allow for richer heterogeneity in the paper.

Under this assumption, we have that the following IV estimand recovers $h_\rho(x)$:

$$\frac{E[\log Y \mid X = x, Z = 1] - E[\log Y \mid X = x, Z = 0]}{E[\log(1-\bar{\tau}) \mid X = x, Z = 1] - E[\log(1-\bar{\tau}) \mid X = x, Z = 0]}.$$

The optimal earnings choice for each individual under full information and tax system d

is then given by

$$Y_i^*(d) = Y_i(d) \left(\frac{1 - \tau_i^*(d)}{1 - \bar{\tau}_i(d)} \right)^{h_\rho(X_i)}.$$

The full-information earnings distribution for $d = 0$ can be computed from the control group, while the full-information earnings distribution for $d = 1$ can be computed from the treated group.

5 Can non-experimental methods recover experimental effects of the tax credit?

Evaluating the causal effects of policies is a central task in empirical economics. Because policies are rarely randomized, researchers generally rely on quasi-experimental methods to construct counterfactual outcomes. Despite the widespread use of non-experimental methods, there is limited direct evidence on whether they recover the causal effects of the policies they are used to evaluate. One important reason is that credible validation requires both an experimental benchmark and sufficiently rich data to reproduce plausible non-experimental evaluations of the same policy. The Norwegian tax-credit experiment provides a unique setting that satisfies both requirements. Random assignment provides an experimental benchmark, while links to population-wide administrative data provide detailed information on individuals, families, firms, earnings, employment, transfers, and consumption over time. We follow the seminal work of LaLonde (1986) and consider the following thought experiment: suppose all individuals of a given group of birth cohorts received the tax credit. What empirical strategies could credibly inform us about the causal effects of the tax credit?

To answer this question, we use the tax credit experiment to assess the performance of several commonly used non-experimental methods across a broad range of outcomes. We consider methods based on selection on observables, panel data, regression extrapolation, eligibility cutoffs, and structural labor-supply models. We aim to answer the following two questions:

Q1: Which methods perform well and why?

We consider the performance of each method in terms of both bias and precision for a number of different outcomes. We then ask why it does or does not perform well. Which identifying assumptions appear to be violated? Does there appear to be selection on unobservables? Is it helpful to use more flexible estimators? Can panel data mitigate concerns with time-varying or time-invariant unobservables? We also ask whether the restrictions from theory used by structural methods increase or decrease the performance of the methods.

Our second question uses the answers to the first question to examine how the methods

may be improved.

Q2: Are there ways to improve methods by relaxing identifying assumptions or using more data?

We first consider relaxing the identifying assumptions that appear to be violated. We derive new partial identification results under the weaker assumptions and consider whether the resulting bounds are (i) non-empty, meaning that they are not rejected by the data, and (ii) informative enough to draw interesting conclusions about the target parameters.

An alternative way to improve the performance of existing estimands is to use the panel dimension of the data to model income processes. A third approach uses machine-learning methods to condition on a richer set of observables.

Methodology

We now introduce and discuss the methods we will consider. For each method and estimand, we derive the parameter it recovers under its associated assumptions. We then show how to use experimental variation to identify the same parameter and obtain an experimental benchmark. This ensures that we measure the bias of each estimator relative to its relevant target parameter. We report the non-experimental estimate, its experimental benchmark, and the difference between them. We also examine how performance varies across populations and specifications. In our case,the counterfactual will be constructed using older birth cohorts. We will study whether the performance of the non-experimental methods depends on the age span between the treated and control cohorts. We then ask whether conventional robustness and specification checks predict which estimates recover the experimental benchmark (Heckman & Hotz, 1989). One type of checks will be placebo estimates based on pre-reform data. We conduct these comparisons for a number of outcomes, including employment, earnings, hours, and selected consumption outcomes.

Selection on observables. The first set of methods relies on selection on observables. We implement propensity-score and nearest-neighbor matching, inverse-probability weighting, and doubly robust estimators that combine weighting with outcome regression (Abadie & Imbens, 2006; Hirano et al., 2003; J. M. Robins et al., 1994; Rosenbaum & Rubin, 1983). These methods compare tax-credit recipients with age-ineligible (here older) individuals who have similar observed characteristics. We vary both the cohort range of the comparison group and the pre-treatment information used to construct the comparison. Because these groups do not overlap in age, the comparison relies on similarities in other observed characteristics and assumptions about differences across age cohorts. This tests whether rich administrative histories can account for systematic differences between eligible and ineligible cohorts.

Panel data methods. The second set of methods relies on changes in outcomes over time. We consider difference-in-differences designs that use outcome changes among age-ineligible cohorts to estimate the counterfactual change for the treated group (Abadie, 2005). We vary the width of the age window and the pre-treatment earnings histories used in the analysis. We examine whether common diagnostics, including tests for pre-trends and permutation tests based on placebo treatment assignments, predict bias (Bertrand et al., 2004; Roth, 2022). We also consider synthetic control methods, which construct a weighted comparison group to match the treated group's pre-treatment outcomes (Abadie et al., 2010). We also consider the performance of distributional difference-in-differences approaches, including changes-in-changes and quantile difference-in-differences (Athey & Imbens, 2006; Callaway et al., 2018). Finally, we study pre–post comparisons and extrapolations based on the treated group's earlier earnings paths (Ashenfelter, 1978; Ashenfelter & Card, 1985).

Regression-based methods. The third set of methods uses regression to extrapolate the untreated outcome relationship from age-ineligible cohorts into the eligible age range. We compare pooled regressions with imputation models estimated only on untreated observations. With heterogeneous effects, the treatment coefficient in a pooled regression need not recover the average effect for the treated. Imputation targets this average directly, but still requires a correctly specified model of untreated outcomes.

Regression discontinuities. The fourth approach uses the age boundaries for eligibility directly. Regression discontinuity compares individuals assigned to the credit near a boundary with age-ineligible individuals on the other side. It identifies a local effect if treated and untreated potential outcomes vary continuously with date of birth at the cutoff (Imbens & Lemieux, 2008). We use local-polynomial estimators with robust bias-corrected inference and data-driven bandwidth selection (Calonico et al., 2014). We estimate the experimental treatment effect at the same cutoff.

Discrete-choice labor-supply models. Finally, we evaluate discrete-choice labor-supply models estimated from observational variation in nonlinear budget sets (Aaberge & Flood, 2008). We estimate preferences using the full untreated population, including the experimental control group and excluding everyone assigned to the tax credit. Including the control group provides observations under the baseline schedule throughout the eligible age range, ensuring overlap in age with the treated group. We use the estimated model to predict behavior under both tax schedules and compare the predicted change in the earnings distribution with the experimental benchmark.

Contribution. By comparing non-experimental estimates with experimental estimates of the same target parameters, we determine which methods work, when they fail, and whether they can be improved. Even if our experimental study follows sixteen cohorts

over five years, this project will also inform the choice of methods to evaluate national policies for individuals within a similar age range.

6 State dependence and dynamic labor supply

This project asks how labor-supply responses to a temporary tax incentive evolve over time and whether they persist after the incentive expires. We focus on workers and study how their hours, earnings, and hourly wages respond during the five-year credit and after its expiration. Static comparisons may miss these effects because workers take time to adjust and because working today can raise wages tomorrow. A gradual phase-in may reflect adjustment costs arising from information, job search, contracting, or work schedules. Additional work experience may also raise future wages and sustain labor-supply effects after the credit expires.

Our analysis will be centered around the following four questions:

Q1: What is the time path of treatment effects, and how do they vary across characteristics like age and education?

We answer this question by estimating how treatment effects vary over the duration of the experiment. We estimate average treatment effects for the full sample and also consider how they vary across a range of characteristics, including age, education, lagged earnings, marriage status, and other socioeconomic outcomes.

Our next question then turns to whether the tax credit has persistent effects.

Q2: Is there evidence of state dependency of the tax credit?

To answer this question, we estimate whether the employment induced by the temporary tax credit persists after the credit expires. Such persistence may reflect state dependence: individuals who are induced to start working because of the tax credit may increase subsequent employment due to accumulated experience, stronger job attachment, or lower costs of continued employment.

Q3: How does state dependency affect the economic interpretation and policy implications of the experimental estimate of the labor supply elasticities?

As emphasized by Keane (2011), state dependency in labor supply may reflect experience effects. The presence of such dynamic effects complicates the interpretation of static labor-supply elasticities. Labor supply may not respond much to large changes in static work incentives because it is mostly determined by future returns to work. To answer the question, we build a dynamic life-cycle model incorporating on-the-job learning so that working today affects productivity in future periods. We then use the data from the

experiment to identify and estimate the model, which allows us to quantify the relative importance of static and dynamic work incentives.

Q4: How large are Frisch elasticities and elasticities of intertemporal substitution?

We use the estimated life-cycle model to compute implied Frisch elasticities and elasticities of intertemporal substitution. The Frisch elasticity plays a key role in the design of tax policy and in business-cycle analysis (Conesa et al., 2009; 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 & Werning, 2020).

We will also discuss whether the time profile of the tax credit effects can be informative evidence on the importance of intertemporal substitution. Since the tax credit disappears in 2031, the are clear incentives to work extra hours in 2030 (when tax is low) and less in 2031.

Methodology and preliminary model

We now present a preliminary version of the model we will use to answer these questions. Building on the labor-supply model in Section 3.1, we develop a life-cycle model with endogenous human capital. In each period, individuals choose consumption, labor supply, and saving. We allow consumption and labor to be nonseparable, with ϕ governing their interaction in utility. Human capital H_t evolves with labor supply L_t, and the hourly wage W_t equals the stock of human capital:

$$H_{t+1} = (1-\delta)H_t + \gamma L_t, \qquad W_t = H_t,$$

where δ is the depreciation rate and γ is the return to work experience (Imai & Keane, 2004; Shaw, 1989). Work therefore yields both current after-tax earnings and a dynamic return through higher future wages.

In observational data, the relationship between current work and future wages is confounded by differences in skills, preferences, and career opportunities. The tax credit generates random variation in work incentives that we use to identify experience effects. To see how, let D denote assignment and $\Delta_D X_t$ the experimental effect on outcome X in period t. Because assignment is independent of initial human capital and preferences, the human-capital equation implies

$$\Delta_D W_{t+1} = (1-\delta)\Delta_D W_t + \gamma \Delta_D L_t.$$

Under this wage process, variation in the experimental effects across periods identifies experience accumulation and depreciation.

These estimates allow us to construct the model-implied full marginal return to work, Ω_t. It combines the current net-of-tax wage with the discounted effect of an additional hour on future after-tax earnings. Under an isoelastic preference specification, experimen-

tal changes in Ω_t, labor supply, and consumption identify the within-period curvature parameters:

$$\Delta_D \log L_t = -\frac{\sigma}{\epsilon}\Delta_D \log C_t + \frac{1}{\epsilon}\Delta_D \log \Omega_t,$$

where σ and ϵ govern curvature in consumption and labor. To recover dynamic elasticities, we also need the degree of nonseparability between consumption and labor. Holding discount preferences and asset returns fixed across assignments, the consumption Euler equation gives

$$\Delta_D \log\left(\frac{C_{t+1}}{C_t}\right) = \frac{\phi}{\sigma}\Delta_D \log\left(\frac{u_{t+1}}{u_t}\right),$$

where u_t is the within-period utility index implied by consumption and labor supply. Together, these moments identify the preference parameters needed to compute two dynamic elasticities at each individual's allocation. The Frisch labor-supply elasticity measures how workers shift labor across periods in response to changes in the full return to work, holding the marginal utility of wealth fixed. The elasticity of intertemporal substitution measures how consumption shifts across periods in response to changes in its intertemporal price. Only in the separable benchmark, $\phi = 0$, do these elasticities reduce to $1/\epsilon$ and $1/\sigma$. With nonseparability, both also depend on ϕ and on the individual's consumption and labor allocation.

Finally, we extend the model to allow changes in hours to be costly. Adjustment costs can delay responses to both the introduction and expiration of the credit. Human-capital accumulation links treatment-induced work to subsequent wage growth. Studying, job changes, and wages together help distinguish delayed adjustment from the returns to accumulated experience. We then use the model to compare temporary and permanent tax changes and to measure how much a static analysis misses by ignoring adjustment and experience effects.

7 Tax credits, parental resources, and child development

Governments often justify transfers to low-income families partly by their benefits for children. Whether in-work tax credits deliver such benefits is unclear. The credits raise household income, but they do so by rewarding work. The credits may therefore also change how much time parents spend with their children, the family's daily routines, who cares for the children, and how much stress parents experience. The net effect on children, and the design of credits that support both work and child development, depend on these responses and on how each affects children.

In this project, we combine the tax credit experiment with several registers, one-to-one skill assessments of 4- to 6-year-olds, and a survey of parents' time use, routines, and beliefs. We use these data to answer the following three questions:

Q1: What is the effect of the tax credit on children's skills, health, and development?

Evidence on how family income affects children is mixed. Studies of tax credits and child benefits find that higher income raises children's test scores and improves their health (Dahl & Lochner, 2012, 2017; Hoynes et al., 2015; Milligan & Stabile, 2011), whereas lottery wins have small effects per dollar (Bulman et al., 2021; Cesarini et al., 2016). In the Baby's First Years trial, a monthly unconditional cash transfer had no detectable overall effect on maternal employment (Sauval et al., 2024) and raised parents' spending on child-specific goods and time in early learning activities, but did not detectably affect children's development at age four (Gennetian et al., 2024; Noble et al., 2025). Randomized evidence on in-work support comes from programs that bundled earnings supplements with work requirements and services (Duncan et al., 2011; Huston et al., 2001).

The tax credit experiment randomizes a pure work incentive in a large population that includes many young mothers and fathers. Most of their children experience the credit in early childhood, a sensitive period for skill formation (Cunha & Heckman, 2007; Heckman & Mosso, 2014). Because early effects often fade and sometimes re-emerge (Bailey et al., 2020), we follow the children from preschool into school.

Q2: How does the tax credit change children's environment?

Much of the literature on family income estimates only its total effect on children, and few studies examine how parents respond (Duncan et al., 2023). In-work credits are likely to change several aspects of children's environment: material resources, the amount and timing of parental time and the family routines it supports, care arrangements; parents' stress and mental health, and family stability. These changes can work in opposite directions, and the net effect on children is therefore ambiguous. Expansions of the US earned income tax credit, for example, reduced mothers' time with children (Bastian & Lochner, 2022) but improved their mental health (Evans & Garthwaite, 2014). Responses may also depend on parents' beliefs about returns to investments in children, which predict how they invest (Attanasio et al., 2022; Cunha et al., 2022). Parents who expect high returns to time may protect time with their children when the credit draws them into work.

Q3: Which channels explain the effects on children?

If income drives the effects, in-work credits and unconditional child benefits may benefit children equally. If parental time, routines, care, or stress also matter, the design of the credit matters for children, including whether it draws parents into full-time or nonstandard-hours work and whether mothers or fathers receive it. Separating the channels is difficult because they tend to move together. Random assignment identifies the effect of the credit on each channel, but not each channel's contribution to the effect on children (Heckman et al., 2013). The few existing studies that separate money and time

rely on strong structural or exclusion assumptions (Agostinelli & Sorrenti, 2022; Del Boca et al., 2014). We complement this with our design, because theory predicts that the credit shifts the channels differently across incentive regions and between mothers and fathers.

Data and methodology

Sample and registers. The child sample consists of all children with at least one parent in the informed sample and a birth date before January 2026. We analyze later births separately, with bounds that allow the credit to affect fertility. The registers contain records of birth outcomes, health care use, prescriptions, child welfare involvement, mapping tests in reading and numeracy (grades 1–3), and national tests in reading, numeracy, and English (grade 5). For parents, the registers contain information about disposable income, employment, hours, mental health, separation, and residential moves, and income and wealth data that let us impute consumption (Eika et al., 2020; see Section 2).

Kindergarten data. No Norwegian register yet records individual kindergarten attendance. The national kindergarten register will record start dates and agreed hours from autumn 2028. Until then, we measure early kindergarten use with cash-for-care payments, which NAV reduces by agreed kindergarten hours for children aged 13–19 months, and with the parent survey.

Direct assessments. In selected urban areas, we invite children aged 4–6 and their parents to local science museums, where the Agder project also assessed children (Rege et al., 2024). Trained testers who do not know the parent's assignment administer a 40-minute tablet battery of mathematics, language, and executive function tests. We construct standardized domain and summary scores. In the Agder project, these scores correlated 0.50–0.67 over 10–19 months and detected treatment effects, notably in mathematics, that persisted into first grade. We invite all eligible children of group A parents and a stratified random sample of group B children, and reassess children we first test at age four when they turn six. If participation differs by assignment, we report bounds (Lee, 2009).

Parent survey. Parents complete a tablet survey at the museum, and we also offer a web version nationwide to all parents of children aged 0–10 in the informed sample, which we coordinate with the survey in Section 4. The survey covers a time diary of the previous day; how often parents read and play with the child; family meals, bedtime, and work schedules; kindergarten hours and informal care; spending on the child; parenting stress and depressive symptoms; and beliefs about returns to parental time and money, which we elicit with hypothetical scenarios (Attanasio et al., 2022).

Experimental effects. For child i in period t, we estimate

$$Y_{it} = \beta D_i + \sigma_{c(i)} + X_i'\delta + u_{it} \tag{1}$$

where D_i indicates that at least one of the child's parents is assigned to the tax credit, $\sigma_{c(i)}$ is a fixed effect for which of the child's parents take part in the experiment and for their randomization strata, and X_i contains predetermined child characteristics. Children with two parents in the experiment are more likely to have a treated parent, so these fixed effects ensure that we compare children with the same probability of treatment. In a second specification, we replace D_i with separate indicators for assignment of only the mother, only the father, and both parents. Few children have two treated parents, so estimates for this group will be imprecise. Because the experiment assigns the credit to parents and siblings share parents, we cluster standard errors two-way by mother and father (Abadie et al., 2023; Cameron et al., 2011). We estimate effects at each age we observe outcomes and by the child's age at exposure, the parent's sex, family structure, and pre-treatment earnings.

Mechanisms. We examine the channels in three steps. First, we estimate equation (1) with each channel index as the outcome. Second, we perform a mediation decomposition following Heckman et al. (2013). We regress child outcomes on assignment and the channel indices and split the total effect into the part each channel explains and a residual. We use factor models to correct for measurement error in the survey measures and test whether the channels' effects differ between those randomly assigned the tax credit or not. The decomposition assumes that unmeasured determinants of child outcomes are unrelated to the measured channels, so we treat the shares as descriptive.

Third, we use differences in how the credit shifts the channels to test and relax this assumption. In the phase-in region of Table 1, the credit raises both income and the return to work, so resources may rise while parental time falls and nonparental care increases. In the phase-out region, income and parental time may rise together. Among couples in which both partners are eligible, the experiment can assign the credit to the mother, the father, both, or neither, which shifts maternal and paternal incentives separately. We estimate effects on the channels and on child outcomes separately for these groups. For example, smaller effects on children per unit of income gain in the phase-in region than in the phase-out region would point to parental time as a channel. We then use interactions between assignment and group as instruments for pairs of channels. This approach identifies the channels' effects if the groups shift the channels in different proportions and each channel has the same effect across groups. We also measure beliefs after assignment and test whether the credit changes them before we use them to explain differences in responses.

8 Fairness, trust, and legitimacy of randomized tax policy

The previous sections study how the tax credit affects economic behavior and what these responses reveal about labor supply and information frictions. The experiment may also have consequences for how people view the tax system and the institutions that implement

it.

The tax-credit experiment creates a tension between two principles of fairness: procedural equality, since everyone has an equal chance of receiving the credit, and horizontal equity, since otherwise similar individuals end up facing different tax liabilities. We study whether assignment to the credit affects trust in public institutions, tax morale, and whether people regard the reform as fair and legitimate. We also examine whether personal experience of the tax lottery affects distributive fairness perspectives more generally, using a consequential impartial-spectator task.

Q1: Does assignment to the tax credit affect institutional trust, stated tax morale, and judgments of the policy's fairness?

We compare individuals who were selected for the credit with those who were not. We first examine trust in Parliament, politicians, the Tax Administration, and NAV. This allows us to assess whether responses are concentrated in the political institutions responsible for the policy and the authority implementing it, or extend to public institutions more broadly. We then examine stated reasons for paying taxes and tolerance of other people's tax evasion. Finally, we ask whether respondents regard equal chances of selection as sufficient to make the experiment fair, and whether this judgment depends on their own assignment. Since we have a panel data structure with judgments both before and after the tax credit assignment, it is possible to study who are updating their judgments in different directions and not only the post treatment means.

Together, these outcomes let us distinguish between isolated dissatisfaction and spillover. A person may regard the policy as unfair while retaining trust in public institutions and a strong commitment to paying taxes. Or their dissatisfaction with the policy may extend to institutional trust and tax morale.

Q2: Does personal experience of the tax lottery affect people's fairness perspectives?

To address this question, respondents make a consequential choice as an impartial third party. Two workers have each received NOK 200 for completing the same assignment. A lottery then determines which of them wins an additional NOK 600 bonus. The respondent is informed about who won the lottery and is given NOK 600 to allocate between the two workers. The respondent can allocate the entire bonus to the lottery winner or choose any of the intermediate allocations offered in the survey. One in every 100 respondents is randomly selected, and that respondent's choice determines the actual allocation of the bonus.

The impartial-spectator design removes the respondent's own material interest from the allocation decision and allows us to use the chosen allocation to measure the fairness perspective applied to the distribution. We compare these choices between respondents assigned and not assigned to the tax-credit to test whether personal experience of the tax lottery affects distributive fairness perspectives.

Methodology and conceptual framework

Experimental effects and the analysis population

Following Section 2, the principal comparison is between groups A and B. Group A was assigned the tax credit, while group B retained the baseline tax schedule. Both groups received an SMS directing them to an information page with personalized information about the tax schedule relevant to them. We use administrative assignment to define treatment. For each outcome Y, we estimate

$$Y_i = \alpha + \beta D_i + \sigma_{s(i)} + X_i'\gamma + \varepsilon_i, \tag{2}$$

where D_i indicates assignment to the credit, $\sigma_{s(i)}$ denotes randomization-stratum fixed effects, and X_i contains prespecified characteristics measured before assignment. The coefficient β measures the effect of being assigned to the tax credit relative to the baseline tax schedule. This total effect of assignment includes effects operating through tax liabilities, incentives, expectations, and information about the assigned tax schedule.

The financial value of the tax credit varies with earnings in the absence of the credit. For example, a person earning around NOK 320,000 would gain roughly NOK 30,000 if selected, whereas a person earning around NOK 656,000 would gain only about NOK 100. We will use this variation to analyze whether the effects of assignment depend on the size of the potential financial gain. We therefore also estimate

$$Y_i = \alpha + \beta D_i + \sum_h \delta_h H_{ih} + \sum_h \theta_h (D_i \times H_{ih}) + \sigma_{s(i)} + X_i'\gamma + \varepsilon_i, \tag{3}$$

where H_{ih} denotes prespecified groups defined by predicted financial exposure to the tax credit, constructed by applying both tax schedules to the same pretreatment administrative information. The interaction coefficients θ_h measure how the effect of assignment varies across exposure groups.

Because survey participation occurs after treatment assignment, we will also check whether participation and response rates differ across treatment groups.

Relation and contribution to existing work

Fairness, tax morale, and institutional trust. A large literature argues that tax compliance depends not only on enforcement and financial incentives, but also on nonpecuniary motivations related to fairness, legitimacy, and trust. Luttmer and Singhal (2014) review this literature. More directly related to our setting, Barth et al. (2013) show in Norwegian data that perceived unfairness in the tax system is associated with greater willingness to justify and engage in tax evasion. Being affected by the policy may also affect trust in public institutions. Using Finland's randomized basic-income experiment, Hirvonen et al. (2025) find that receiving basic income increased trust in parliament and politicians. This literature motivates asking whether the randomly assigned tax treatment affects

judgments of the policy itself, institutional trust, and stated tax morale.

Procedural fairness and distributive fairness. An experimental literature distinguishes the fairness of a procedure from the fairness of the resulting allocation. Bolton et al. (2005) show that allocation procedures matter for fairness judgments, while Cappelen et al. (2013) study the tension between equal ex ante opportunities and unequal outcomes generated by luck. The latter study also illustrates the use of impartial spectators to elicit fairness views without a direct material interest in the allocation. This provides the conceptual basis for our distinction between judgments of the tax experiment's equal-chance procedure and the fairness perspective revealed in the consequential allocation task.

Personal experience and fairness perspectives. Our second question relates to a more recent literature asking whether people's own economic experiences shape the fairness principles they apply to others. Cassar and Klein (2019) randomly expose individuals to success or failure generated either by performance or by luck and show that this experience subsequently affects distributive choices over other people's incomes. Blouin et al. (2025) similarly show that randomly generated differences in own earnings affect subsequent redistributive choices concerning others.

Our setting brings these strands together in a large-scale public-policy experiment. Random assignment creates economically meaningful differences in tax treatment, allowing us to study effects on the perceived fairness and legitimacy of the policy, trust and tax morale, and the fairness perspective respondents apply in a consequential impartial-spectator task.

9 A review of other planned projects

9.1 Education

A temporary tax credit may also affect educational choices, through the opportunity cost of time spent studying rather than working. This project asks whether a temporary change in work incentives affects only when people acquire education, or also how much education they end up with. Because the credit is conditional on labor income, it is worth little to a full-time student. In contrast, the credit is worth up to NOK 27,500 a year to someone working instead of studying, so we expect it to affect enrollment, drop-out and delayed entry rather than study intensity. The *Lånekassen* income threshold for full-year grant recipients falls inside the phase-in region of Table 1, so the credit lowers the marginal tax rate on both sides of the threshold and leaves the incentive to earn past it unchanged. We extend the labor-supply model developed in Section 3.1 to a life-cycle setting in which time is allocated between work and education (Becker, 1965). We link assignment status to the National Education Database and estimate effects separately for those enrolled and those not enrolled at randomization. We follow outcomes over 2026–2035 to separate

delayed from forgone education. The outcomes and directional hypotheses by incentive region are pre-specified in a separate pre-analysis plan.

9.2 Fertility

This project asks whether the credit changes the number and the timing of births, and whether any response is driven by the opportunity cost of time or by income and liquidity. Parental benefits are part of the credit's calculation base, so parental leave at full coverage does not reduce it, and a lower average tax rate raises the affordability of children. At the same time, a parent on leave cannot increase hours, so where the credit improves work incentives, acting on them means postponing childbearing. The predicted sign is therefore unambiguous in the plateau and phase-out regions of Table 1 and ambiguous in the phase-in region. We extend the labor-supply model developed in Section 3.1 to a life-cycle setting in which time is allocated between work and children (Becker, 1965). We link assignment to birth records from *Folkeregisteret* and estimate effects on births, entry into parenthood, and higher-order births over 2026–2035 to separate changes in timing from changes in completed fertility. The outcomes and directional hypotheses by incentive region are pre-specified in a separate pre-analysis plan.

9.3 Consumption and saving responses

This project asks how a temporary tax reduction affects the allocation of consumption over time. We extend the labor-supply model developed in Section 3.1 to allow for intertemporal decisions. Random assignment identifies effects on consumption, saving, debt, and other balance-sheet outcomes. Our detailed data allow us to study how tax-credit recipients allocate the additional disposable income generated by the reduction in taxes across consumption, saving, and debt repayment. We estimate consumption and saving responses relative to the gain in disposable income. Because the credit also changes work incentives, these ratios need not equal marginal propensities to consume or save out of an exogenous income increase. To interpret them, we combine the intertemporal model with the labor-supply estimates from Section 3, separating mechanical tax relief from income changes due to labor-supply responses. We study how these responses vary with predetermined characteristics such as pre-treatment income, liquidity, debt, assets, homeownership, and predicted exposure to the credit.

The Nets data allow for a more detailed understanding of when the additional resources are spent and which goods they are spent on. In addition, detailed data on real-estate and financial transactions allow us to examine whether relatively small changes in disposable income relax financial constraints and thereby induce larger, lumpy adjustments, such as home purchases, debt repayment, or financial-asset transactions. The project therefore shows whether recipients smooth the temporary gain over time or use it to make larger balance-sheet adjustments.

9.4 Household labor supply

This project studies household labor supply. When partners share resources, their work decisions depend on the household budget, individual work incentives, and preferences over consumption and leisure. Changes in one partner's tax incentives may therefore affect the other partner's labor supply. Correlated wages and preferences make identifying these responses difficult (Blundell et al., 2016).

We use the tax credit experiment to address these difficulties: Among couples where both partners are eligible, the credit may be assigned to the wife, the husband, both, or neither. This experimental variation lets us study how each partner's labor supply responds to changes in their own and their partner's work incentives.

We link partners in the population registers and estimate effects of each assignment configuration on both partners' employment, hours, and earnings. We extend the non-parametric labor-supply model in Section 3.1 to allow preferences to be nonseparable in consumption and both partners' labor supply, with unrestricted heterogeneity across households. We aim to go beyond uncompensated own- and cross-price elasticities to learn about income effects and compensated price responses. We derive identification results for these parameters under alternative economic assumptions.

We also assess the restrictions imposed by unitary and collective models of household decision-making. We use the estimated responses to study counterfactual tax credits, including how their effects depend on which partner receives them. This provides useful evidence on the consequences of individual versus joint taxation (Golosov & Krasikov, 2025).

9.5 Health

This project studies how the tax credit affects treatment for physical and mental health problems. Higher disposable income may improve health by reducing financial strain and financing health investments. Changes in employment affect exposure to physical and psychological demands, social and workplace contact, and time use, all of which can have positive or negative effects on health. The net health effect is therefore ambiguous.

We link assignment to population-wide administrative data on health-care utilization, diagnoses, prescriptions, sickness absence and benefit take-up. We start by estimating the net effect of the tax credit on health-care visits, diagnoses and prescription use. We then ask to what extent these changes can be explained by treatment-seeking, employment, and benefit incentives. It is important to consider both health and employment responses jointly, as working more may have direct effects on health-care visits through the need to certify sickness absence and changing the time available for treatment, without any changes in underlying health. To probe deeper into the mechanisms, we exploit large linked health-surveys with standardized health measurements.

9.6 Crime

This project asks whether the tax credit affects criminal activity by easing financial strain and raising the returns to legal work. We link random assignment to administrative crime records and combine the experimental estimates with a model of legal work and crime to distinguish between these channels.

Increased financial strain typically increases crime for affected populations (Deshpande & Mueller-Smith, 2022; Dustmann et al., 2024; Tuttle, 2019). By contrast, no statistically significant effect of lottery wealth on criminal convictions were found in Sweden (Cesarini et al., 2023). Much less is known about the effects of raising the returns to legal work. The fact that the tax credit changes the returns to declared work in a systematic way along the income distribution, ref. Table 1, allows us to ask whether the crime response follows the returns to legal work or the resources. As these mechanisms relate closely the employment effects of the tax credit, we build on the labor-supply model in Section 3.1 to develop a model of legal work and crime to distinguish between these channels.

The target group for the tax credit is young adults, who are the age group with the highest crime rates. The credit is also aimed at low earners, which is a segment of the population with relatively high crime rates. The tax credit trial is thus well-suited to study the effects on crime. Using the long individual-level Norwegian crime records permits pre-assignment criminal history to be built for every cohort in the trial, providing a basis heterogeneity. Prior work attachment will help determine who has access to adjust work hours. Property crime is income-motivated and thus the primary crime outcome family. In follow-up work, we will study effects on undeclared work.

References

Aaberge, R., & Flood, L. (2008). *Evaluation of an in-work tax credit reform in sweden: Effects on labor supply and welfare participation of single mothers* (IZA Discussion Paper No. 3736). IZA.

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, *72*(1), 1–19.

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, *105*(490), 493–505.

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267.

Afriat, S. N. (1967). The construction of a utility function from expenditure data. *International Economic Review*, *8*(1), 67–77.

Agostinelli, F., & Sorrenti, G. (2022). *Money vs. time: Family income, maternal labor supply, and child development* (HCEO Working Paper No. 2018-017, revised October 2022). University of Chicago.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495.

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.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, *24*(2), 3–30.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, *60*(1), 47–57.

Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, *67*(4), 648–660.

Athey, S., & Imbens, G. W. (2006). Identification and inference in nonlinear Difference-in-Differences models. *Econometrica*, *74*(2), 431–497.

Attanasio, O., Boneva, T., & Rauh, C. (2022). Parental beliefs about returns to different types of investments in school children. *Journal of Human Resources*, *57*(6), 1789–1825.

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, *21*(2), 55–97.

Barth, E., Cappelen, A. W., & Ognedal, T. (2013). Fair tax evasion. *Nordic Journal of Political Economy*, *38*, 1–16.

Bastian, J., & Lochner, L. (2022). The earned income tax credit and maternal time use: More time working and less time with kids? *Journal of Labor Economics*, *40*(3), 573–611.

Becker, G. S. (1965). A theory of the allocation of time. *The Economic Journal*, *75*(299), 493–517.

Bensnes, S., Hernæs, Ø., Håvarstein, M., Markussen, S., Mogstad, M., Raaum, O., & Torsvik, G. (2026, October). *Forsøksordning med arbeidsfradrag – rapport 1: Gjennomføringen* (Memo). Frischsenteret.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust Difference-in-Differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Bierbrauer, F. J., Boyer, P. C., & Hansen, E. (2023). Pareto-improving tax reforms and the earned income tax credit. *Econometrica*, *91*(3), 1077–1103.

Bitler, M. P., Gelbach, J. B., & Hoynes, H. W. (2006). What mean impacts miss: Distributional effects of welfare reform experiments. *American Economic Review*, *96*(4), 988–1012.

Blouin, A., Mani, A., Mukand, S. W., & Sgroi, D. (2025). The meritocratic illusion: Inequality and the cognitive basis of redistribution. *Oxford Economic Papers*, *77*(4), 1128–1147.

Blundell, R., Browning, M., Cherchye, L., Crawford, I., De Rock, B., & Vermeulen, F. (2015). Sharp for SARP: Nonparametric bounds on counterfactual demands. *American Economic Journal: Microeconomics*, *7*(1), 43–60.

Blundell, R., Browning, M., & Crawford, I. (2008). Best nonparametric bounds on demand responses. *Econometrica*, *76*(6), 1227–1262.

Blundell, R., Pistaferri, L., & Saporta-Eksten, I. (2016). Consumption inequality and family labor supply. *American Economic Review*, *106*(2), 387–435.

Bolton, G. E., Brandts, J., & Ockenfels, A. (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal*, *115*(506), 1054–1076.

Bulman, G., Fairlie, R., Goodman, S., & Isen, A. (2021). Parental resources and college attendance: Evidence from lottery wins. *American Economic Review*, *111*(4), 1201–1240.

Callaway, B., Li, T., & Oka, T. (2018). Quantile treatment effects in Difference-in-Differences models under dependence restrictions and with only two time periods. *Journal of Econometrics*, *206*(2), 395–413.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249.

Cappelen, A. W., Konow, J., Sørensen, E. Ø., & Tungodden, B. (2013). Just luck: An experimental study of risk-taking and fairness. *American Economic Review*, *103*(4), 1398–1413.

Cassar, L., & Klein, A. H. (2019). A matter of perspective: How failure shapes distributive preferences. *Management Science*, *65*(11), 5050–5064.

Cesarini, D., Lindqvist, E., Östling, R., & Schroeder, C. (2023). *Does wealth inhibit criminal behavior? evidence from swedish lottery winners and their children* (Working Paper No. 31962). National Bureau of Economic Research.

Cesarini, D., Lindqvist, E., Östling, R., & Wallace, B. (2016). Wealth, health, and child development: Evidence from administrative data on Swedish lottery players. *The Quarterly Journal of Economics*, *131*(2), 687–738.

Cherchye, L., Demuynck, T., & De Rock, B. (2014). Revealed preference analysis for convex rationalizations on nonlinear budget sets. *Journal of Economic Theory*, *152*, 224–236.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica*, *80*(3), 969–1018.

Chetty, R., Friedman, J. N., & Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review*, *103*(7), 2683–2721.

Chetty, R., & Saez, E. (2013). Teaching the tax code: Earnings responses to an experiment with EITC recipients. *American Economic Journal: Applied Economics*, *5*(1), 1–31.

Conesa, J. C., Kitao, S., & Krueger, D. (2009). Taxing capital? not a bad idea after all! *American Economic Review*, *99*(1), 25–48.
Cunha, F., Elo, I., & Culhane, J. (2022). Maternal subjective expectations about the technology of skill formation predict investments in children one year later. *Journal of Econometrics*, *231*(1), 3–32.
Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review*, *97*(2), 31–47.
Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*, *102*(5), 1927–1956.
Dahl, G. B., & Lochner, L. (2017). The impact of family income on child achievement: Evidence from the earned income tax credit: Reply. *American Economic Review*, *107*(2), 629–631.
Del Boca, D., Flinn, C., & Wiswall, M. (2014). Household choices and child development. *The Review of Economic Studies*, *81*(1), 137–185.
Deshpande, M., & Mueller-Smith, M. (2022). Does welfare prevent crime? the criminal justice outcomes of youth removed from SSI. *The Quarterly Journal of Economics*, *137*(4), 2263–2307.
Duncan, G., Kalil, A., Mogstad, M., & Rege, M. (2023). Investing in early childhood development in preschool and at home. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 1–91, Vol. 6). Elsevier.
Duncan, G. J., Morris, P. A., & Rodrigues, C. (2011). Does money really matter? estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology*, *47*(5), 1263–1279.
Dur, R., Harms, J., & Non, A. (2026). *Tax misperceptions and labor supply: A randomized information experiment* (IZA Discussion Paper No. 18870). IZA Institute of Labor Economics.
Dustmann, C., Landersø, R., & Andersen, L. H. (2024). Refugee benefit cuts. *American Economic Journal: Economic Policy*, *16*(2), 406–441.
Dutz, D., & Håvarstein, M. (2026). *A revealed preference analysis of labor supply and income taxation* [Unpublished manuscript].
Edmark, K., Liang, C.-Y., Mörk, E., & Selin, H. (2016). The swedish earned income tax credit: Did it increase employment? *FinanzArchiv: Public Finance Analysis*, *72*(4), 475–503.
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.
Eissa, N., & Hoynes, H. W. (2004). Taxes and the labor market participation of married couples: The earned income tax credit. *Journal of Public Economics*, *88*(9–10), 1931–1958.
Eissa, N., & Liebman, J. B. (1996). Labor supply response to the earned income tax credit. *Quarterly Journal of Economics*, *111*(2), 605–637.
Evans, W. N., & Garthwaite, C. L. (2014). Giving mom a break: The impact of higher EITC payments on maternal health. *American Economic Journal: Economic Policy*, *6*(2), 258–290.
Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276.
Forges, F., & Minelli, E. (2009). Afriat's theorem for general budget sets. *Journal of Economic Theory*, *144*(1), 135–145.
Gennetian, L. A., Duncan, G. J., Fox, N. A., Halpern-Meekin, S., Magnuson, K., Noble, K. G., & Yoshikawa, H. (2024). Effects of a monthly unconditional cash transfer starting at birth on family investments among US families with low income. *Nature Human Behaviour*, *8*(8), 1514–1529.
Gideon, M. (2017). Do individuals perceive income tax rates correctly? *Public Finance Review*, *45*(1), 97–117.
Golosov, M., & Krasikov, I. (2025). The optimal taxation of couples. *The Quarterly Journal of Economics*, *140*(3), 2163–2211.
Hall, R. E. (1988). Intertemporal substitution in consumption. *Journal of Political Economy*, *96*(2), 339–357.
Hausman, J. A., & Wise, D. A. (1979). Attrition bias in experimental and panel data: The gary income maintenance experiment. *Econometrica*, *47*(2), 455–473.
Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *American Economic Review*, *83*(2), 116–121.
Heckman, J. J., & Hotz, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *Journal of the American Statistical Association*, *84*(408), 862–874.
Heckman, J. J., & Mosso, S. (2014). The economics of human development and social mobility. *Annual Review of Economics*, *6*, 689–733.
Heckman, J. J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086.
Heckman, J. J., Smith, J. A., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, *64*(4), 487–535.
Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*(4), 1161–1189.

Hirvonen, S., Schafer, J., & Tukiainen, J. (2025). Policy feedback and voter turnout: Evidence from the Finnish basic income experiment. *American Journal of Political Science*, *69*(4), 1388–1405.
Hoderlein, S., & Stoye, J. (2014). Revealed preferences in a heterogeneous population. *The Review of Economics and Statistics*, *96*(2), 197–213.
Hoynes, H., Miller, D., & Simon, D. (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy*, *7*(1), 172–211.
Huston, A. C., Duncan, G. J., Granger, R., Bos, J., McLoyd, V., Mistry, R., Crosby, D., Gibson, C., Magnuson, K., Romich, J., & Ventura, A. (2001). Work-based antipoverty programs for parents can enhance the school performance and social behavior of children. *Child Development*, *72*(1), 318–336.
Imai, S., & Keane, M. P. (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review*, *45*(2), 601–641.
Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.
Jacquet, L., Lehmann, E., & Van der Linden, B. (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, *148*(5), 1770–1805.
Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, *49*(4), 961–1075.
Kitamura, Y., & Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, *86*(6), 1883–1909.
Kleven, H. (2024). The EITC and the extensive margin: A reappraisal. *Journal of Public Economics*, *236*, 105135.
Kleven, H. J., & Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from pakistan. *The Quarterly Journal of Economics*, *128*(2), 669–723.
Kline, P., & Tartari, M. (2016). Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *American Economic Review*, *106*(4), 972–1014.
Kostøl, A. R., & Myhre, A. S. (2021). Labor supply responses to learning the tax and benefit schedule. *American Economic Review*, *111*(11), 3733–3766.
LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*(4), 604–620.
Laroque, G. (2005). Income maintenance and labor force participation. *Econometrica*, *73*(2), 341–376.
Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, *76*(3), 1071–1102.
Luttmer, E. F. P., & Singhal, M. (2014). Tax morale. *Journal of Economic Perspectives*, *28*(4), 149–168.
Manski, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics*, *5*(1), 145–174.
Meyer, B. D., & Rosenbaum, D. T. (2001). Welfare, the earned income tax credit, and the labor supply of single mothers. *Quarterly Journal of Economics*, *116*(3), 1063–1114.
Milligan, K., & Stabile, M. (2011). Do child tax benefits affect the well-being of children? evidence from Canadian child benefit expansions. *American Economic Journal: Economic Policy*, *3*(3), 175–205.
Noble, K., Magnuson, K., Duncan, G., Gennetian, L. A., Yoshikawa, H., Fox, N. A., Halpern-Meekin, S., Troller-Renfree, S., Han, S., Egan-Dailey, S., Nelson, T. D., Mize Nelson, J., Black, S., Georgieff, M., & Karhson, D. (2025). *The effect of a monthly unconditional cash transfer on children's development at four years of age: A randomized controlled trial in the U.S.* (NBER Working Paper No. 33844). National Bureau of Economic Research.
Prescott, E. C. (2006). Nobel lecture: The transformation of macroeconomic policy and research. *Journal of Political Economy*, *114*(2), 203–235.
Rees-Jones, A., & Taubinsky, D. (2020). Measuring "schmeduling". *The Review of Economic Studies*, *87*(5), 2399–2438.
Rege, M., Størksen, I., Solli, I. F., Kalil, A., McClelland, M. M., ten Braak, D., Lenes, R., Lunde, S., Breive, S., Carlsen, M., Erfjord, I., & Hundeland, P. S. (2024). The effects of a structured curriculum on preschool effectiveness: A field experiment. *Journal of Human Resources*, *59*(2), 576–603.
Riddell, C., & Riddell, W. C. (2024). Welfare versus work under a negative income tax: Evidence from the gary, seattle, denver, and manitoba income maintenance experiments. *Journal of Labor Economics*, *42*(2), 427–467.
Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*(427), 846–866.
Robins, P. K. (1985). A comparison of the labor supply findings from the four negative income tax experiments. *Journal of Human Resources*, *20*(4), 567–582.
Robins, P. K., & West, R. W. (1986). Sample attrition and labor supply response in experimental panel data: A study of alternative correction procedures. *Journal of Business & Economic Statistics*, *4*(3), 329–338.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322.
Russell, T. M. (2021). Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, *39*(2), 532–546.
Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, *68*(1), 205–229.
Saez, E. (2002). Optimal income transfer programs: Intensive versus extensive labor supply responses. *Quarterly Journal of Economics*, *117*(3), 1039–1073.
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.
Sauval, M., Duncan, G. J., Gennetian, L. A., Magnuson, K. A., Fox, N. A., Noble, K. G., & Yoshikawa, H. (2024). Unconditional cash transfers and maternal employment: Evidence from the Baby's First Years study. *Journal of Public Economics*, *236*, 105159.
Shaw, K. L. (1989). Life-cycle labor supply with human capital accumulation. *International Economic Review*, *30*(2), 431–456.
Straub, L., & Werning, I. (2020). Positive long-run capital taxation: Chamley–Judd revisited. *American Economic Review*, *110*(1), 86–119.
Torgovitsky, A. (2019). Partial identification by extending subdistributions. *Quantitative Economics*, *10*(1), 105–144.
Tuttle, C. (2019). Snapping back: Food stamp bans and criminal recidivism. *American Economic Journal: Economic Policy*, *11*(2), 301–327.