Hours Mismatch and Annual Taxes*

Andreas R. Kostøl† Andreas S. Myhre ‡ Matthew Merkle§ Mark Whitmeyer¶

This Version: July 22, 2025
Preliminary and incomplete

Abstract: How does a mismatch between desired and offered hours affect labor supply responses to annual taxes? In this paper, we show theoretically that hours constraints can induce workers to stop working as their cumulative taxable income creeps into a new tax bracket mid-year. We develop an empirical approach that identifies this intertemporal extensive margin labor supply response non-parametrically from monthly employment data, and apply it to annual kinks in the Norwegian tax and transfer system. Overall, we find the elasticity of labor supply is 0.8 among marginally attached workers, but less than 0.1 for high-earners. We relate this "missing mass of work" to "excess mass of year-end earnings" at the kink, 60% of workers are subject to indivisible hours. We use this fraction to quantify the role of hours mismatch in explaining divergence between macro- and micro evidence on intertemporal participation elasticities, and find that in the limiting case – in which all workers are hours constrained, the intertemporal elasticity of substitution would be 1.3. Our findings suggest that hours constrained and marginally attached workers can account for changes in the aggregate workforce over the business cycle.

Keywords: labor supply; incentives; hours constraints; tax and transfer system.
JEL codes: H20; H31; H55; J22; J26

*This project received financial support from the Research Council of Norway project 345156. This project has benefited greatly from conversations with Richard Blundell and suggestions from participants at the IFN Stockholm Conference and 3rd Oslo Labor-Public Economics Workshop. All errors are our own.

†BI Norwegian Business School; IZA. E-mail: andreas.r.kostol@bi.no
‡Statistics Norway. E-mail: andreassmyh@gmail.com
§WP Carey School of Business, ASU. E-mail: matthew.merkle@asu.edu
¶Department of Economics, Arizona State University; E-mail: mark.whitmeyer@gmail.com

1 Introduction

A growing body of micro-level evidence shows that workers are often mismatched with respect to their desired hours of work (e.g., Lachowska *et al.*, 2023). More flexible hours schedules are often associated with lower pay and constitute an important contribution to the gender wage gap (Goldin, 2014). Since the early work by Hansen (1985) and Rogerson (1988), rigid hours offered by employers has also been considered central to understanding the large fluctuations in employment over the business cycle. However, macroeconomic models are often calibrated using participation elasticities that remain inconsistent with micro-level evidence, and there is little, if any, evidence on how work incentives interact with hours mismatch.[1] In this paper, we provide a new approach to identifying the prevalence of hours constraints and how they shape the response to annual taxes.

To make progress on this question, we first demonstrate that the presence of hours constraints, when combined with annual taxation, generates a distinct participation response. Unlike the traditional static labor supply response to annual tax incentives, hours constraints introduce a dynamic participation decision—whether to continue working or stop as cumulative income crosses into a higher tax bracket. This participation decision involves quitting or temporarily pausing work until the tax resets in January of the following year, where pausing work is preferred when the utility cost of over-working exceeds the cost of quitting.[2] We show that this participation response around tax kinks can only be rationalized by hours mismatch, and not by other behavioral frictions, such as learning about complex incentives.

We develop an empirical approach that allows us to identify the participation response from monthly employment data. The central element to our approach is the participation probability function (PPF), capturing the likelihood of working the next month as a function of year-to-date cumulative income. In theory, this function has a rectangular shape as cumulative income creeps into a new tax bracket from the left. Under this assumption, the counterfactual PPF in the absence of the kink—or higher tax bracket—is obtained by

[1] Keane (2011) and Saez *et al.* (2012) survey the literature on dynamic labor supply decisions and the elasticity of taxable income respectively. A growing number of studies in public economics show that responses to incentives are often dampened by various kinds of frictions (e.g., Chetty, 2012 and Kleven & Waseem, 2013). Saez (2010) argued that these frictions explain why employees often do not respond to tax incentives, and Chetty *et al.* (2013) show that hours constraints can rationalize the divergence between Hicksian macro and micro elasticities, but not the intertemporal elasticity of labor supply: "Even accounting for indivisible labor, micro studies do not support representative- agent macro models that generate Frisch elasticities above 1" (Chetty *et al.*, 2013, p. 4).

[2] Quitting costs also captures the cost of returning to work in the next tax period, or at a later date. Of course, pausing work may impose high career costs to full time workers (Kleven *et al.*, 2025), but to some workers, ignoring the tax incentive can be costlier than quitting.

simple extrapolation. The participation response is identified nonparametrically by comparing the change in the observed PPF to the right of the bracket to the counterfactual PPF (from the left).

We apply this framework to the Norwegian setting, which offers a unique combination of monthly earnings data and a wide range of tax brackets within the tax and transfer system. We begin by exploring the responses to large kinks among part-time workers who receive partial benefits from the disability insurance (DI) system. While the sample is restricted to marginally attached workers, the main advantage of this setting is that the cost of ignoring the incentive is unusually high, with a marginal tax at the kink of 50 percent. We find that a temporary reduction in the net after tax income of 62 percent leads to a fall in the participation rate by 53 percent, yielding an implied participation elasticity of 0.8. This result is robust across a wide set of specifications and holds across part-time workers who return to the same or a new employer the next calendar year, and those who have left the workforce when the tax resets the following year.

Whereas hours constraints limit the ability to fully optimize, our approach allows us to further decompose year-end bunching into an intensive vs. extensive margin response, and to estimate the overall fraction of workers who are hours constrained. By relating the amount of "missing mass of work" to the "excess mass in year-end earnings", we can derive the share of constrained "compliers"—i.e., workers who find it optimal to locate at or close to the kink. We find that while the missing mass of work accounts for about 2 percent of the part-time employed welfare recipients, the missing mass accounts for 60 percent of "compliers." Next, we use this fraction to quantify the role of hours mismatch in explaining divergence between macro- and micro evidence on intertemporal participation elasticities. In the limiting case, in which all marginally attached workers are hours constrained, the intertemporal participation elasticity would be at least 1.3.

As highlighted by Chetty (2012), theory suggests inertia when the utility cost of ignoring the incentive is low relative to the adjustment cost. To test these predictions, we draw on the full spectrum of kinks in the tax system, which range from marginal tax changes of 3 to 25 percentage points and span the 10th to the 90th percentile of the earnings distribution. Consistent with the theory, we find evidence of extensive margin responses at the lowest bracket but much weaker evidence at higher brackets. Moving from zero to 25 percent at the first tax bracket induces a labor supply response of 5 percent, with an implied elasticity of 0.2. At the higher end of the distribution, we find that the participation elasticity is different from zero but smaller than 0.1 for one kink, while estimates for other kinks are imprecise.

The participation elasticities we estimate are greatest among marginally attached work-

ers. One explanation to why we find large extensive margin responses is that the population has low search costs, making labor market entry and exit relatively easy. Indeed, many workers quit and return to the same firm the following year, suggesting that while hours constraints are rigid, and that firms can accommodate entry and exit. Relatedly, career concerns are likely less important, as the majority of workers hold part-time positions or are nearing retirement age. Consistent with this notion, we find that participation elasticity is smaller among workers in larger firms—where job flexibility may be greater—and among younger workers—for whom career concerns play a more significant role.

Our paper is related to research on implications of optimization frictions for labor supply decisions. Saez (2010) discusses the role of hours frictions as a reason to why workers do not bunch, and Chetty *et al.* (2011) point to aggregate factors, such as collective wage bargaining, that may attenuate the estimated micro response to incentives. Chetty (2012) formalizes the idea of optimization frictions, and Kleven & Waseem (2013) show how notches and dominated regions in budget sets can be used to identify the overall prevalence of optimization frictions.[3] Relative to this research, our approach nests the standard bunching approach of Saez (2010), in which people make a single labor supply decision.[4] Our approach allows a decomposition of the bunching into intensive and extensive margins. We show that hours constraints do not attenuate the annual elasticity of earnings in setting in which the quitting costs are high.

Our work also contributes to research on responses to changes in work incentives (e.g., Moffitt, 1992, Krueger & Meyer, 2002, Saez *et al.*, 2012 and Kleven & Waseem, 2013). While it is well-established that labor supply elasticities differ widely across groups or countries, understanding the sources remains an active research field (see, e.g., Krueger & Meyer, 2002 and Chetty, 2012). Our paper contributes to this line of research by unpacking the role of hours frictions in shaping the elasticity of labor supply.[5] In related work, Weathers & Hemmeter (2011) and Ruh & Staubli (2019) find large labor supply responses to return-to-work programs in the US and Austrian disability insurance (DI) systems, in which

[3] A number of studies have extended the bunching framework to allow for adjustment costs; see, for example, Gelber *et al.* (2019), Ruh & Staubli (2019) and Zaresani (2020).

[4] Our paper is methodologically related to Saez (2003)—who uses gradual changes in annual earnings relative to locations of kinks to estimate taxable income elasticity. Similarly, our approach to estimating the counterfactual density is related to Kopczuk & Munroe (2015)—who use data only from the left of a kink, which is not affected by the introduction of a tax.

[5] A large literature studies how optimization frictions affect decisions, including labor supply responses to welfare reform with adjustment frictions (Kline & Tartari, 2016), over-withholding of tax liabilities (Jones, 2012), information provision and take-up of the earned income tax credit (Chetty & Saez, 2013; Bhargava & Manoli, 2015), and labor supply of old-age retirees (Liebman & Luttmer, 2012). Hastings & Weinstein (2008) and Jensen (2010) study how information and perceptions about the returns to schooling affect behavior. Chetty *et al.* (2009) and Finkelstein (2009) document attenuated responses to non-salient taxes in consumer spending.

benefits are taxed monthly. In contrast, Zaresani (2020) finds small earnings responses to financial incentives in the Canadian DI system, in which benefits are taxed annually. Our findings suggest that labor market frictions interact with the safety net with important consequences for welfare and inequality.

Our approach relates to existing research on macro and micro evidence on the elasticity of intertemporal substitution. Chetty *et al.* (2013) show that reduced-form estimates of this extensive margin response tend to reflect a combination of the share of workers whose wages are close to their reservation wage and labor market frictions such as adjustment costs and hours constraints. In our setting, interpreting the participation response as an intertemporal extensive margin substitution elasticity requires an assumption of forward-looking behavior. When the change in effective wages is anticipated, there is no wealth effect—only a temporary change in net after-tax income. Our study offers evidence that could help rationalize Frisch elasticities of 2, at least for a population of marginally attached workers, who are most sensitive to changes in take-home wage rates.

Finally, our findings have policy implications. While tax incentives typically kick in at year-end, welfare systems often calculate benefit offsets at the monthly level. The US SSDI program, for example, phases out DI payments at a monthly frequency. Our evidence speaks to the policy proposal known as the "$1 for $2 offset" that aims to stimulate increased exit from the US SSDI program by allowing recipients to keep $1 of benefits for every $2 of labor earnings above the substantial gainful activity (SGA) threshold. Our evidence indicates that improving the work incentives of marginalized workers, such as disabled workers with some residual work capacity, can have large effects on part-time work and the aggregate workforce.

The remainder of the paper is organized as follows. Section 2 describes a model where people make repeated labor supply decisions over the year. Section 3 describes the policy environment and the data. Section 4 presents our main findings and Section 5 concludes.

2 Theory and Empirical Methodology

We first develop a theory to study extensive margin responses to taxes under hours constraints. The main conclusion is that individuals facing hours constraints may find it optimal to stop working mid-year, while other frictions do not generate participation effects. We propose estimating a conditional participation probability to analyze the effects of taxes on the extensive margin.

2.1 A Dynamic Model of Hours Mismatch and Year-End Incentives

Individuals face a tax period with 12 sub-periods, $t = 1, \ldots, 12$, discounting the future according to δ. Throughout, a worker's per-period wage is w, and the cumulative labor income is denoted $y_{t-1} = w \sum_{i=1}^{t-1} h_i$. Following Diamond (1980) and Saez (2002), we study a general tax and transfer system with negative taxes at zero labor income, and denote non-labor income $x > 0$. We assume that there is no tax withholding, so that this cash benefit is distributed equally per month, always guaranteeing some income for consumption, allowing us to omit savings decisions from the baseline specification.[6]

Each period, given labor supply choice h_t and cumulative income y_{t-1}, a worker chooses how much to consume, c_t, subject to the constraint, $c_t \leq h_t w \left(1 - \tau \left(y_{t-1}\right)\right) + x$, where $\tau \left(y_{t-1}\right)$ gives the tax schedule (effective tax paid on cumulative income up to last period y_{t-1}). The tax (or equivalently, the benefit reduction) takes the following single-kinked form:

$$\tau \left(y_{t-1}\right) = \begin{cases} 0, & \text{if} \quad y_{t-1} \leq K \\ \tau, & \text{if} \quad K < y_{t-1} \end{cases},$$

where $0 < \tau < 1$ is a scalar tax rate, and $K > 0$ is the location of a kink in the tax schedule. We assume the tax rate is perfectly understood. In the appendix, we show that there is no qualitative difference between a tax and a benefit reduction for the purposes of our identification result.

A worker draws her disutility of labor θ from some distribution, and we assume for simplicity it is drawn before making work and consumption choices. A worker of type θ's per-period utility, net of labor costs is

$$u \left(c_t\right) - \theta D \left(h_t\right),$$

where $u \left(\cdot\right)$ is strictly increasing, twice continuously differentiable, and concave. The disutility of work $D(\cdot)$ is also increasing, twice continuously differentiable, and strictly convex.[7] Without savings, a worker will consume her entire income in each period; given cumulative income y_t, her per-period payoff as a function of labor supply h_t is

$$u \left(h_t w \left(1 - \tau \left(y_t\right)\right) + x\right) - \theta D \left(h_t\right).$$

We analyze this problem for two types: a (i) flexible and (ii) hours constrained worker.

[6] In effect, this means that over- and under-withholding is administered the following year.

[7] The leading example of such a cost is the isoelastic utility form, $\frac{\theta}{1+\frac{1}{e}} h_t^{1+\frac{1}{e}}$, where e is the elasticity of labor supply with respect to net after-tax earnings.

2.1.1 Analysis for a Flexible Worker

We begin by examining the problem of a worker who can freely choose their hours of labor supply. In the simplest version of the model, savings are not allowed. However, as long as a worker freely determines their hours each period, the problem exhibits no time inconsistency and is effectively static, reducing to the choice of a sequence of per-period labor supply decisions $\{h_t\}_{t=1}^{12}$.

In the absence of tax withholding, if total desired earnings exceed the kink, there exists a period s ($s \in \{0, \ldots, 12\}$) that is the last period before earnings strictly exceed the kink. In period $s+1$, a portion of that period's earnings becomes taxable, but only a partial increase in tax may occur, as cumulative earnings up to period s may still be below the kink threshold. From period $s+2$ onward, all income is subject to taxation. If total earnings exceed the kink starting from period 1 onward, $s = 0$. If total earnings never hit strictly above the kink, $s = 12$.

We assume that for all types θ, the condition $wu'(x) - \theta D'(0) > 0$ holds, ensuring that all workers find it optimal to work in at least one period within the year. The worker chooses $\{h_t\}_{t=1}^{12}$ to maximize

$$\overbrace{\sum_{t=1}^{s} \delta^{t-1} \left(u\left(h_t w + x\right) - \theta D\left(h_t\right)\right)}^{\text{periods strictly below kink}} + \overbrace{\sum_{t=s+2}^{12} \delta^{t-1} \left(u\left(h_t w\left(1-\tau\right) + x\right) - \theta D\left(h_t\right)\right)}^{\text{periods strictly above kink}} + \overbrace{\delta^{s} \left(u\left(w h_{s+1}\left(1-\tau\right) + w\tau(K - \sum_{i=1}^{s} h_i) + x\right) - \theta D\left(h_{s+1}\right)\right)}^{\text{period crossing the kink}} \tag{1}$$

This problem can be solved in two steps. First, fix $s \in \{0, \ldots, 12\}$, then optimize Expression 1 subject to the constraints that

$$w \sum_{t=1}^{s} h_t \leq K, \quad \text{and} \quad w \sum_{t=1}^{s+1} h_t > K.$$

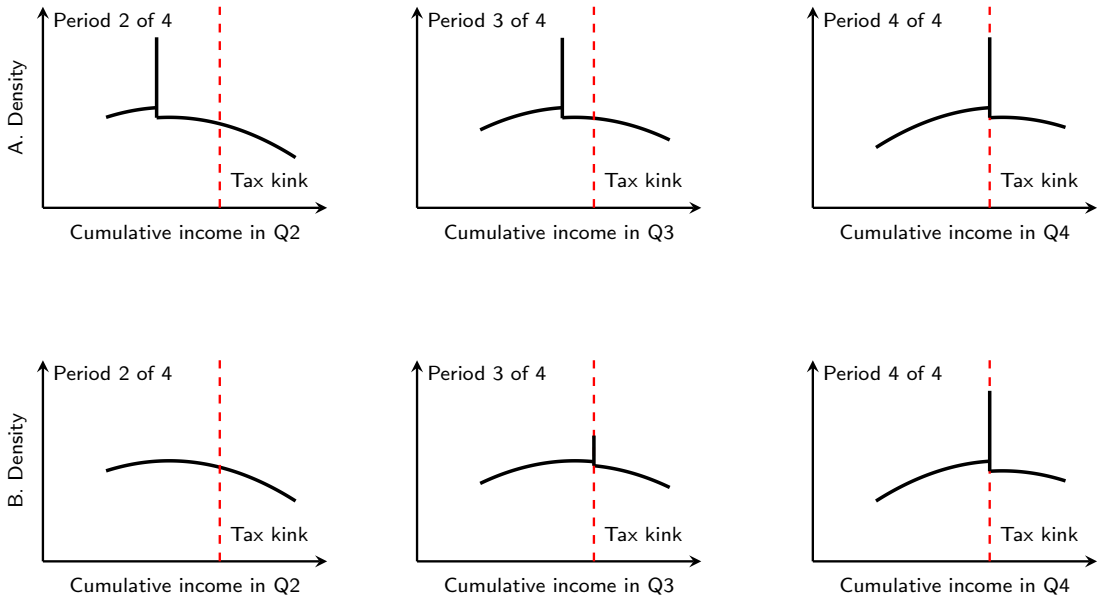
Second, given the solution to this program, optimize the choice of s. Fixing s, the regularity of the objective ensures that an optimizer of Expression 1 exists. As s can take only finitely many values, a solution to the grand (two-stage) problem exists. A special case of this problem is that in which a worker finds it optimal to have her annual earnings be exactly equal to the kink point K. In that case, the worker solves a standard "cake-eating" problem.

Proposition 1. *If a flexible worker's marginal disutility of work at* 0, $D'(0)$, *is sufficiently low, it is uniquely optimal to work in every period.*

That is, a flexible worker's labor supply h_t must be strictly positive in every period t, unless $D'(0)$ is sufficiently large. If, for instance, the disutility of work at 0 hours is 0, it is optimal to work in every period. We provide a proof of this proposition in the Appendix, and show in the next section that it also holds when allowing for savings.

Assuming that there exists a smooth distribution of the disutility of work, Panel A of Figure 1 illustrates the dynamics of the cumulative earnings distribution with $T = 4$. As workers find it optimal to smooth work across periods, some workers will also find it optimal to bunch their work hours, such that the per period work decisions accumulates to the annual kink, as seen in the rightmost figure. This set of compliers—i.e., those who find it optimal to bunch—is constant in each period, illustrated by the fixed height of the excess mass, which moves rightward throughout the year.

Figure 1: Illustration of Cumulative Earnings Distribution



Notes: Panel A illustrates a possible distribution of cumulative quarterly earnings with perfect control of hours worked, and Panel B illustrates it for the hours constrained type.

2.1.2 Analysis for an Hours-Constrained Worker

Turning to an hours-constrained worker, the key difference is that the worker faces an exogenously specified (from the worker's perspective) amount of work per period $h_t = h >$

0. In principle, to overcome the hours constraint, a worker could exert effort at some cost to adjust hours–to search for a different job or negotiate different hours with the current employers, for instance. When this cost is sufficiently high, the constrained worker's only labor supply choice each period is whether to continue working or to quit . We present a simple model, without considering the search decision, as a way to overcome the adjustment cost. While it is possible to pause work, we only consider quitting. Consequently the worker simply decides the period to stop work $f \in \{0, 1, \ldots, 12\}$, where $f = 12$ means the worker does not quit. The quitting decision also comes at a cost $\psi > 0$, which captures its negative effect on the worker's income going into the next year (and other negative consequences).

To avoid trivialities, we assume that $wh < K$ and $u(hw + x) - \theta D(h) > u(x)$, ensuring that a worker finds it optimal to work at least one period. We say that a worker stops working strictly below the kink if $f < 12$ and $w(f + 1)h \leq K$. That is, a worker stops working strictly below the kink if she could work at least one more period before her total income exceeded the kink. First, we note that even with hours constraints, a worker never stops working strictly below the kink.

Remark 1. It is never optimal for a worker to stop working strictly below the kink.

Proof. We simply show that working one more period strictly improves upon stopping before the kink. The difference in annual utilities from a worker stopping at $f + 1 \leq 12$ and f has the same sign as $u(hw + x) - \theta D(h) - u(x)$, which is strictly positive, by assumption. □

In contrast to the flexible hours case, a worker facing hours constraints may either over- or under-work. Importantly, an hours-constrained worker may find it optimal to stop working mid-year.

Remark 2. It is optimal for a worker to stop working mid-year if $hw \cdot 11 > K$ and $u(hw(1 - \tau) + x) - \theta D(h) \leq u(x) - \psi$.

Proof. We simply show that stopping in period 11 strictly improves upon not stopping. But this is just the stated condition. □

Panel B of Figure 1 illustrates the evolution of cumulative earnings for hours-constrained workers. It shows that hours constraints—in combination with an annual kink—induce some workers to stop working mid-year as their cumulative earnings reach the kink. In that case, the worker finds it optimal to stop working because her take-home wage, $w(1 - \tau)$, is too low compared to her disutility of working, $\theta D(h)$. The figure highlights

that while year-end earnings may be identical across workers, within-year behavior is key to understanding the role of hours mismatch in shaping labor supply responses.

2.1.3 Extensions

In our model, we made a number of simplifications. In particular, we abstract away from saving, we assume that workers are properly specified (they understand the tax schedule), we do not allow for random shocks, we model benefit withholding as a tax, and impose an additively-separable form for a worker's utility $u\left(c_t, h_t; \theta\right) = u\left(c_t\right) - \theta D\left(h_t\right)$. Now we briefly explain that our result differentiating flexible versus hours-constrained workers is robust to relaxing these assumptions, deferring the discussion of benefit withholding and the more general utility $u\left(c_t, h_t; \theta\right)$ to the appendix. Suppose that workers may save and, for the sake of contraposition, that one finds it optimal to stop working mid-year. Without loss of generality, we assume that her stopping point is after her annual income has exceeded K. Given a stock of savings $\gamma \geq 0$, and letting $V(\cdot)$ denote the worker's continuation value from carrying the stock of savings into the next year, the worker solves

$$\max_{\{c_t,\}_{t=1}^m, \{f_t\}_{t=1}^m} \sum_{t=1}^{m} \delta^{t-1} u\left(c_t\right) + V(f_m),$$

subject to $c_1 + f_1 \leq \gamma + b$ and $c_t + f_t \leq b + f_{t-1}$ for all $2 \leq t \leq m$. Let c_1^* be part of a solution to this program. As it is optimal not to work, it must be that working a little bit and consuming all of it–leaving future behavior unchanged–is suboptimal, i.e.,

$$w(1-\tau)u'\left(c_1^*\right) - \theta D'(0) \leq 0,$$

which is false as long as $D'(0)$ is sufficiently close to 0.

Extending the framework to include preference shocks, shifting the disutility of working throughout the year—for example, due to the presence of young children or health shocks—is straightforward. The main assumption we must invoke is that preference shocks evolve smoothly over time, and are not correlated with the timing that cumulative earnings enters the vicinity of kinks or notches in the budget constraints.

Finally, we extend the framework to allow for misspecified workers, assuming that a fraction $\alpha \in (0,1)$ of workers mistakenly believe the kink in the benefits schedule is instead at some $\hat{K} \neq K$. Crucially, a misspecified worker is unaware of this misperception. Consequently, early in the year (for low t), such a worker behaves similarly to a correctly specified worker, except that she optimizes based on the incorrect kink. Suppose a worker learns the correct specification upon exceeding the kink: at that point, she observes a

change in the tax, realizes that she is above the kink, and learns its true location. Upon being informed of the correct kink—regardless of how the information is obtained—the worker reoptimizes.

The key question is whether reoptimization can induce a worker to stop working. To address this, we consider the problem of a worker whose misspecification is corrected. One possibility is that the worker becomes aware of her misspecification before exceeding the kink. In this case, her problem is identical to that of a correctly specified worker, with the only difference being that she has a strictly positive endowment (her accumulated earnings) and a shorter time horizon (as fewer months remain in the year). The other possibility is that the worker becomes aware of her misspecification only after exceeding the kink. Would she then stop working mid-year? The answer is no, as long as the marginal disutility of work at zero hours is close to zero. Moreover, following the logic of Lemma 1 in Appendix B, it is never optimal for a misspecified worker to stop strictly before reaching the kink.

2.2 Empirical Methodology and Identification

The basic idea behind the empirical approach is that the width of the response region is determined by the severity of hours mismatch. The theory establishes a relationship between the distance from start-to-current period cumulative income and the kink. The closer an individual is to the kink, the more likely we are to observe a decline in participation. We operationalize this idea using a participation probability density (PPF) function, which depends on cumulative income, $p_{e,t}$, capturing the fraction who work in cumulative income bin e at time period t. We denote the lower and upper bound of the response region L and U, and calculate the conditional participation probability density (PPF).

We illustrate the response region by the shaded triangle that appears in the third and fourth period in Figure 2. The counterfactual PPF is denoted $\tilde{p}_{e,t}$, and captures the participation in the absence of the kink; the counterfactual PPF is represented by the dotted line above the shaded triangle in Figure 2. Using the counterfactual PPF, the participation response at the kink is identified from the following expression

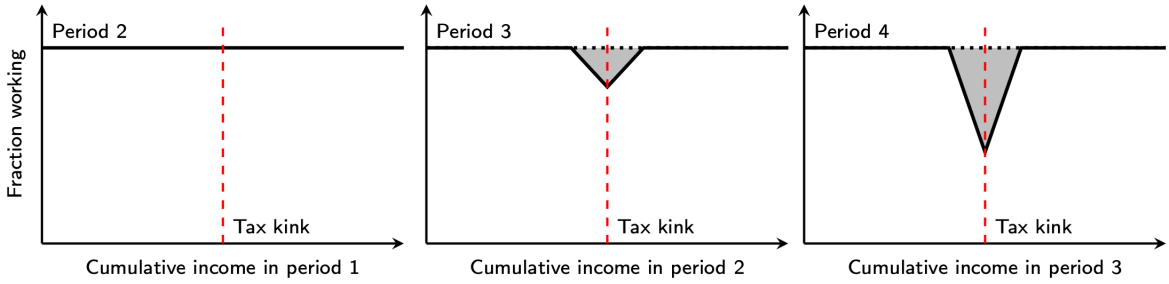
$$\Delta p = E[\tilde{p}_{e,t} - p_{e,t} | e = K],$$

which in the illustration is equal to the vertical distance between the solid line (actual PPF) and the dotted line (counterfactual PPF). Similarly, the total amount of missing mass m_t of work can be estimated from the response region as the following expression,

$$m_t = \frac{1}{\tilde{p}_{K,t}} \sum_{e=L}^{U} \left(\tilde{p}_{e,t} - p_{e,t} \right),$$

which is equal to the integral over the shaded triangle divided by the *counterfactual mass* at the kink $\tilde{p}_K$. The total number of earnings bins in the response region is given by $(U - L)$.

Figure 2: Illustration of Missing Mass



Notes: This figure displays an example of the actual and counterfactual participation probability density (PPF) functions in the bunching region around the tax kink when $T = 4$. The actual PPF is represented by the solid black line and the counterfactual PPF is represented by the dashed black line. The shaded triangle denotes the response region and the total area of the shaded triangle—the total difference between the actual and counterfactual PPFs—is the mass of missing work, m_t.

To identify the PPF, we use a histogram estimator with equally sized bins. The counterfactual PPF is estimated by first fitting a constant and a linear slope using data to the left of the response region. While the theory suggests this line is flat, our approach allows for an upward or downward-sloping trend. We then extrapolate a straight line from the lower bound of the response region to the kink or the upper bound. We also allow a more flexible counterfactual following the polynomial approach from Chetty *et al.* (2011).

2.2.1 Identifying the Constrained Share of Compliers

Next, we want to understand the distribution of constrained optimizers, relative to the overall excess mass at the end of the year. The idea of the bunching approach is similar to the dynamic framework we have presented, but assumes individuals make one decision of how many hours to supply in a year. As in our framework, there exists a smooth distribution of the disutility parameter θ, which ensures a smooth distribution of earnings. Only individuals with initial labor earnings in the range $[K, K + \Delta z]$, will find it optimal to bunch exactly at (or near) the kink in response to the introduction of a new tax bracket with marginal tax rate τ. This behavior generates an excess mass of earnings among compliers at the end of the tax period, denoted b.

We can decompose the role of the participation margin in shaping the excess mass. We do so by relating the cumulative missing mass to the excess mass in the following way

$$E[\alpha_t \mid e = K] = \frac{m_T}{b_T} \cdot \frac{\sum_{e=L}^{U} \tilde{M}_{e,t}}{\tilde{N}_{K,T}},$$

where $\tilde{M}$ is the counterfactual frequency in a given cumulative earnings bin, and $\tilde{N}_{K,T}$ is the counterfactual frequency of individuals locating their period T earnings at the kink.

2.2.2 Implementation

We implement the rectangular approach and the polynomial approach to estimate the missing mass, and the polynomial approach for the excess mass.[8] We calculate standard errors by generating earnings distributions by 500 random draws of the sample with replacement. We then re-estimate the parameters of interest within each sample, and define the standard error as the standard deviation of the distribution of interest, and perform block-bootstrap with replacement over individuals, keeping all observations of each individual that we resample to account for potential serial correlation in the error terms.

2.3 A Reduced Form Approximation of the Participation Elasticity

The participation elasticity we estimate will, in general, be attenuated by the fact that many workers are not facing an hours constraint, and instead respond to incentives by adjusting the intensive margin. The elasticity is simply defined as

$$\varepsilon = \frac{\triangle p/\tilde{p}}{\triangle(1-PTR)/(1-PTR)} \tag{3}$$

Naturally, this elasticity is attenuated by those who respond to taxes primarily by changing hours. To correct this elasticity for the prevalence of hours constraints, we use

[8] We follow standard practice and assume a flexible polynomial fitted to the histogram of earnings—excluding the bunching region—which pins down the counterfactual histogram height at the kink $\hat{n}_K$ (see, e.g., Chetty *et al.*, 2011). Specifically, we fit the following function

$$n_j = \sum_{i=0}^{p} \beta_i \left(z_j\right)^i + \sum_{k=z^L}^{z^U} \mathbf{1}\left(z_j = k\right) + \varepsilon_j, \tag{2}$$

where n_j is the number of workers in earnings bin j, z_j is the earnings level, z^L is the lower bound of the bunching region—below K, when people face optimization frictions or make forecast errors—and z^U is the upper bound of the bunching region. Both upper and lower bounds can typically be verified by visual inspection. Flexibly controlling for earnings in this range with the dummies $\gamma_k \mathbf{1}\left(z_j = k\right)$ allows us to estimate the polynomial function without considering the behavioral response to the kink.

our estimated share (α) in the denominator to identify the limiting case in which all workers are hours constrained. Another important issue is that this elasticity only captures hours constraints that are binding from below – i.e., those who would like fewer hours than their employer can offer. Our approach is unable to speak to the role of having multiple employers to reach a desired amount of consumption and leisure.

3 Policy Environment and Data

This section describes the policy environment, data, and main findings.

3.1 Policy Environment

Our empirical analysis draws on two attractive features of the Norwegian labor market. The first feature is the policy environment, which includes a tax system with multiple tax brackets at the lower and upper ends of the earnings distribution. Moreover, the safety net allows for a combination of cash payments to eligible workers and earned income, subject to sharp benefit offsets. Chetty (2012) shows that large changes in incentives are more likely to induce a response as the welfare cost of ignoring the incentive is high. For this reason, we begin by studying Disability Insurance (DI), which is the largest transfer program besides retirement pension, and allows recipients to work while receiving DI cash benefits. The key advantage of the DI program is that it features a large kink, where marginal tax rates change by approximately 50 percent at an income level equivalent to an average of 10 hours per week. [9]

The benefit administration is similar to other countries, in that cash is paid on a monthly basis. Similar to the US system, beneficiaries can earn up to an annual substantial gainful activity (SGA) threshold without losing any benefits. In Norway, this level is equal to $8,000 annually, where earnings up to this point are subject to income taxation. While the US Social Security DI system has a phase-out based on a monthly threshold, the Norwegian system uses an annual earnings threshold (see, e.g., Kostøl & Mogstad, 2014 for details). When cumulative earnings exceed the annual SGA threshold, benefits are re-

[9] The level of DI benefits received is based on a worker's previous earnings and calculates a worker's average indexed annual earnings (AIE). Past wages are indexed to present wages using a deflator equal to the average wage growth in the economy. For awards before 2015, years with the lowest earnings are excluded, and the proportion of income replaced falls with the level of previous earnings. After 2015, the benefits are calculated based on the average earnings the last five years before disability onset, which is determined by a physician. The system can also grant partial awards, allowing beneficiaries to use a larger fraction of their residual work capacity.

duced by about \$0.66 for every dollar in gross earnings.[10] Accounting for income taxes, the kink still more than doubles the effective marginal tax from, on average, 0.35 to 0.78, where the individual slope change varies with the replacement rate. Recipients remain eligible for DI benefits until their earnings exceed a threshold set to 80 percent of a weighted average of their past earnings, after which benefits are terminated.

3.2 Data

The second key feature is the monthly paycheck data. What is unique to this data is that it covers every employment relationship—including information about earnings, contracted hours, hourly wages, and bonuses—and allows us to distinguish between strategic reporting of income from self-employment and other types of behavioral responses. Each employer reports earnings and cash benefits paid, and the data is continuously updated as new files from employers are received. We link this data to monthly files on DI recipients provided by the SSA—which provides additional information on the level of DI benefits, pre-DI earnings (AIE), the date of award, whether the award was for partial or full DI benefits; and the age, gender, and household information of recipients. Finally, we link this data to data provided by Statistics Norway which contains information on educational attainment, and annual earnings and cash transfers from other welfare programs based on tax registers.

Table 1 presents summary statistics for the overall workforce (Panel A) and the population of DI recipients (Panel B). The table shows that the majority of earnings come from regular pay, except for DI recipients, whose earnings primarily consist of variable income in addition to cash payments from the DI system. Additionally, welfare recipients tend to be older and less educated compared to the overall workforce.

An important benefit of the Norwegian administrative data is that all income components, except income from self-employment, are third-party reported (e.g. by employers and financial intermediaries). We use the employment data only, to avoid having to deal with self-employment income which is only reported annually. Moreover, the coverage and reliability are rated as exceptional by international quality assessments (see e.g. Atkinson *et al.*, 1995). Since administrative data are a matter of public record, there is no

[10] There are two ways in which the benefit phase-out is administered. If individuals expect to earn above the SGA amount, they can report their year-end earnings expectations directly to the SSA, which leads to immediate adjustments to benefits. This is equivalent to tax withholding. The other is by ex post adjustments, allowing for implicit borrowing with repayment the following calendar year. Since 2015, employers report income monthly to the SSA, and the DI system starts phasing out benefits once recipients' cumulative monthly earnings exceed the annual SGA amount. Before 2015, these ex post adjustments were made after the DI program received information from the tax authorities, which was typically around August in the following calendar year.

attrition due to non-response or non-consent by individuals or firms, and individuals can only exit these datasets due to natural attrition (i.e., death or out-migration).

Table 1: Summary Statistics

	A. Workforce		B. DI Recipients	
	Mean	SD	Mean	SD
Column:	(1)	(2)	(3)	(4)
Age	41.4	(13.7)	50.0	(11.6)
Fraction female	.47		.54	
Monthly earnings ($)	5,225	(3,210)	971	(1,247)
Regular pay ($)	3,771	(3,232)	328	(855)
Variable pay ($)	1,048	(1,812)	498	(836)
Fraction regular pay	.71		.25	
Fraction hourly pay	.26		.43	
Fraction fee-based/other pay	.03		.33	
Contracted hours per week	32.8	(12.7)	12.7	(15.3)
Fraction full time	.68		.19	
Fraction part time	.28		.49	
Fraction no contracted hours	.03		.32	
Years of schooling	11.9	(4.3)	10.2	(2.7)
Observations <individuals>	27,449,916	2,871,511	655,369	47,009

Notes: This table provides summary statistics of the overall population and DI recipients. Panel A reports statistics for the overall (non-disabled) workforce below age 66. Panel B reports statistics for DI recipients with DI receipt in 2015 and full DI award prior to 2015. Outcome variables are measured in 2016 and 2017. Age, years of schooling, cohabitant status, and number of children are measured in 2015. All other covariates are either pre-determined or constant over time. Earnings and DI benefits are measured in 2016 dollars (NOK/USD = 7.5).

4 Evidence

This section presents the main results on participation effects and the prevalence of hours constraints using kinks from the safety net and the tax system.

4.1 Responses to the Benefit Offset Kink

We begin our analysis with the DI system due to the salience of benefit offset incentives. Our sample includes all DI recipients with some labor income who receive full benefits at the beginning of 2015.[11] We study responses to the kink at $8,000, and measure outcomes during 2015–2017.

[11] These restrictions are due to a lack of information on the exact location of the kink among the partially disabled and because recipients awarded benefits after this date faced a different set of work incentives. Additionally, we exclude awards from after 2015—as they were subject to a kink of about $5,000—and those aged above 66—due to eligibility for old-age retirement benefits beginning at age 67.

4.1.1 Participation Responses

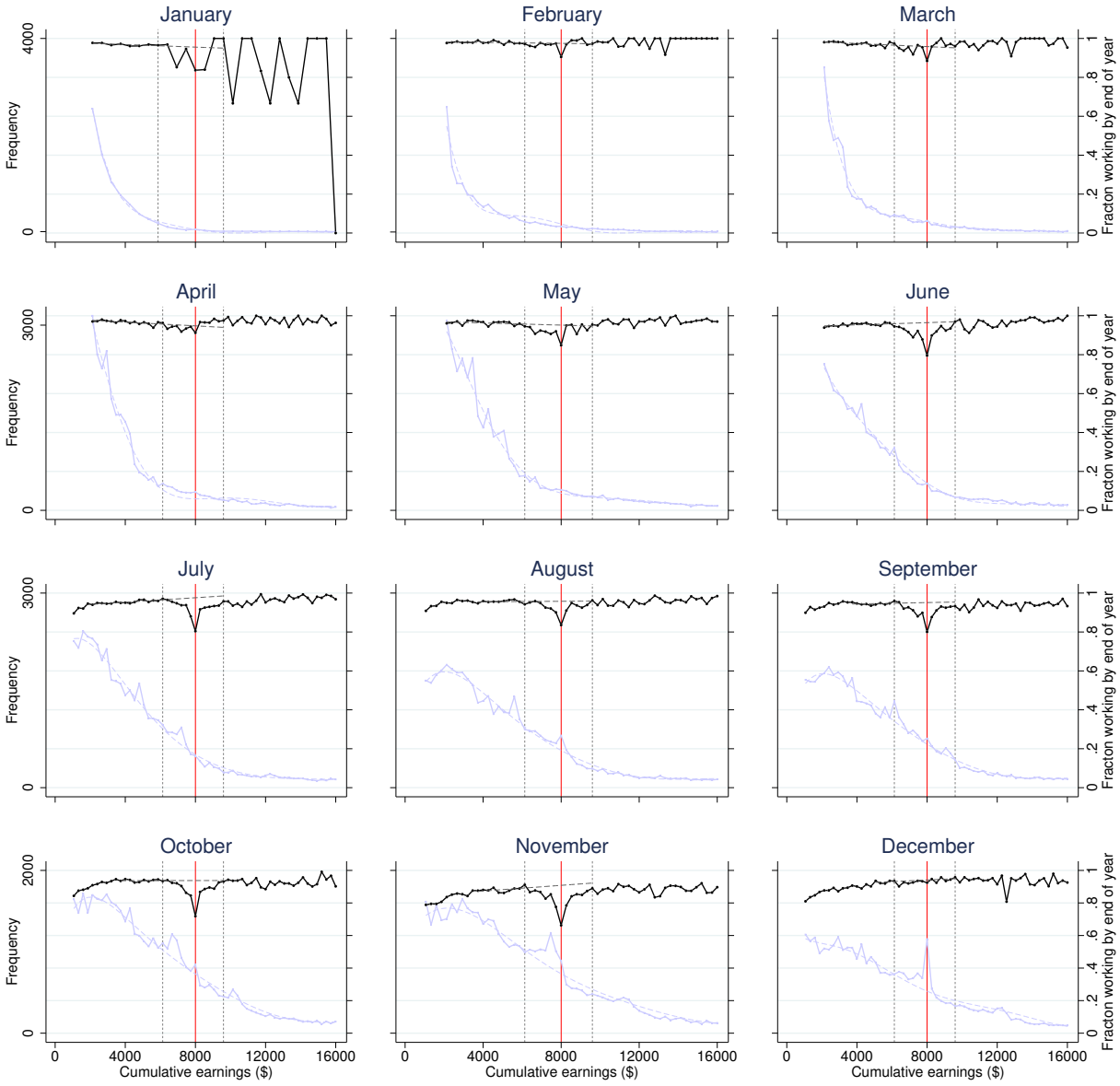
A virtue of our approach is that the conditional work probabilities can be easily plotted. According to the theory, missing mass around the kink is evidence that hours frictions constrain individual responses to tax incentives. We plot the work decision in all months of the year as a function of cumulative earnings up to the previous month in Figure 3. We compute the probability to work as a function of year-to-date cumulative earnings around the kink for each calendar month, and plot these probabilities with solid black lines and the counterfactuals with dashed lines. We estimate the counterfactual work probabilities for each bin using a linear extrapolation of the observations to the left of the excluded region. We also plot the cumulative earnings distributions using solid blue lines, and the respective counterfactual distribution with a dashed line. Our baseline specification estimates the counterfactual earnings distributions using a seventh-degree polynomial, omitting bins in the excluded region, as indicated by the vertical dashed lines.

In the first month, very few individuals have earnings close to the annual $8,000 kink. The PPF estimates are noisy during the first few months, particularly for cumulative earnings above the kink. However, we see clear signs of extensive margin responses in the later months: The fraction of recipients who continue working is significantly lower for those whose cumulative earnings are around the kink compared to recipients below (and above) the kink. This indicates that the work incentives created by the kink induce some recipients to stop working as their cumulative earnings reach the $8,000 mark.

Figure 3 shows that this response becomes more pronounced later in the calendar year; Appendix Figure A.1 plots the size of the missing mass of work by calendar month. We estimate that few individuals respond at the extensive margin early in the year. The missing mass is fairly low and insignificant until April but increases in May, where we estimate that 126 individuals do not work due to the work incentives created by the kink. The size of the response continues to rise in subsequent months, reaching 776 missing workers—or a missing mass of 1.13—in November.

Our estimates of extensive margin responses—i.e., the fraction of individuals whose earnings are exactly $8,000 when they stop working—range from 0.04 to 0.12 during January to April and are borderline statistically significant. The response strengthens over the year, reaching approximately 0.25 in November, as shown graphically in Appendix Figure A.4. By contrast, we find no evidence of extensive margin responses in the next calendar year based on year-end earnings in December, with point estimates remaining insignificant. This placebo test confirms that the temporary incentive to stop working is driving the participation result rather than other unobserved factors.

Figure 3: Missing Mass by Month



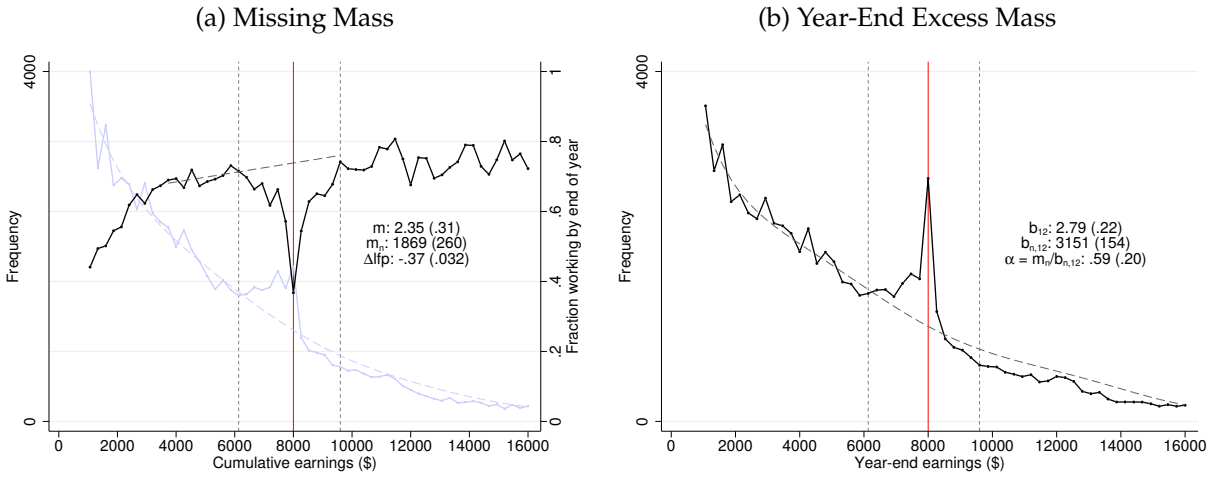
Notes: These figures display the observed cumulative earnings distributions for the sample of DI recipients and fitted counterfactual distributions using a 7th degree polynomial as described in equation 2 for each calendar month (left axes). The figures further display the observed fractions of recipients who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$267 = 2,000 NOK) and the estimated counterfactual probabilities ($L\hat{F}P$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axes). Missing work (m) is defined as the total weighted difference between LFP and $L\hat{F}P$ in the excluded region around the kink relative to $L\hat{F}P$ at the kink, and is weighted by the estimated counterfactual frequency. The excluded region is indicated by the dashed vertical lines. Similarly, missing man-months (m_n) is defined as the total weighted difference between LFP and $L\hat{F}P$ in the excluded region, weighted by the estimated counterfactual frequency. $\triangle LFP$ is defined as the difference between LFP and $L\hat{F}P$ at the \$8,000 kink. For the full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

To aggregating these monthly participation responses to an annual number, we must deal with two issues. The first is that some may quit and start work multiple times within

a calendar year. Second, by pooling all observations we would double-count individuals who stop working at the kink in June, and do not start working until the next calendar year. A potential consequence of double counting is that our measure of missing mass is upward biased. To address this issue, we use only the last observation for each individual (up until November) if they work multiple months during the calendar year. This can be interpreted as the individual's final optimization decision within the year, and ensures that each individual is counted only once.

Figure 4 presents estimates of responses using the last observation only. Panel (a) displays the observed and estimated counterfactual cumulative earnings distributions, indicated by the solid blue and dashed lines, respectively. The observed and estimated counterfactual fraction of recipients who work by the end of the calendar year for each earnings bin are represented by the solid and dashed black lines, respectively.

Figure 4: Missing Mass of Work and Excess Mass of Earnings at Disability Insurance Benefit Kink



Notes: Figure A displays the observed cumulative earnings distribution for the sample of DI recipients and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). The figure further displays the observed fraction of recipients who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$267 = 2,000 NOK) and the estimated counterfactual probability ($L\hat{F}P$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axis). Missing work (m) is defined as the total weighted difference between LFP and $L\hat{F}P$ in the excluded region around the kink relative to $L\hat{F}P$ at the kink, and is weighted by the estimated counterfactual frequency. The excluded region is indicated by the dashed vertical lines. Similarly, missing man-months (m_n) is defined as the total weighted difference between LFP and $L\hat{F}P$ in the excluded region, weighted by the estimated counterfactual frequency. $\triangle LFP$ is defined as the difference between LFP and $L\hat{F}P$ at the \$8,000 kink. Figure B displays the observed year-end earnings distribution and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). Year-end bunching (b_{12}) is defined as the total difference between the observed and counterfactual distribution in the excluded region around the kink relative to the counterfactual at the kink. Similarly, the number of year-end bunchers ($b_{n,12}$) is defined as the total difference between the observed and counterfactual distribution in the excluded region. For the full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Visually, there are clear signs of individuals responding at the extensive margin as their cumulative earnings reach the kink. Individuals with cumulative earnings around

the kink appear far less likely to continue working compared to those below (and above) the kink. We estimate a missing mass of 2.35, meaning that the total amount of missing work in the excluded region around the kink corresponds to 2.35 times the counterfactual work probability at the kink. Similarly, we estimate that a total of 1,869 individuals do not continue working due to the work incentives at the kink.

We estimate a drop in the labor force participation rate of 0.37 compared to the counterfactual work probability at the kink, which is approximately 0.74. This corresponds to a 50% decrease in the labor force participation rate for individuals with cumulative earnings at the kink. These estimates are highly significant and clearly indicate large extensive margin responses due to the increased marginal tax.

Table 2 reports the estimated labor force participation responses and implied participation elasticity. We estimate a labor force participation response of approximately 0.37 percentage points at the kink. Compared to our preferred estimate of the counterfactual probability of continued work at the kink—0.74, using the linear extrapolation method—this corresponds to a 50 percent drop in the labor force participation rate for individuals whose cumulative earnings amount to $8,000 (the kink point). This translates to a participation elasticity of approximately 0.80 with respect to the (net-of) participation tax rate.

In row B, we estimate the counterfactual labor force participation rate by fitting a flexible polynomial to the observed fraction of working individuals while omitting observations in the excluded region. The estimates are highly robust to the choice of estimation method, with the estimated labor force participation response at the kink being 0.35 and the participation elasticity being 0.78.

In row C, we examine whether our responses may be driven by individuals who quit working altogether and who may not be subject to hours constraints. To address this, we estimate labor force participation responses using only individuals who continue working in the next calendar year. Since our data span through 2017, we restrict this specification to observations from 2015–2016. Compared to our baseline labor force participation estimate of 0.39, the estimate is higher—around 0.44—when including only individuals who also work the following year. This suggests that our results are not driven by individuals who exit the labor market entirely. The response is very similar for individuals who continue working at the same firm as in the baseline year but somewhat lower for those who switch to another firm the following year. Conversely, our estimated responses are lower for individuals who do not work the next year—as shown in the last row–with the estimated labor force participation response being approximately 0.20. However, this appears to be driven by a much lower overall labor force participation rate of around 0.32.

Table 2: Estimates of the Labor Force Participation Elasticity

	Extensive margin response ($\triangle p$)		Labor force participation ($\tilde{p}$)		Participation elasticity (ϵ)		Obs
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A: Work by end of year (2015-2017)	-.37	(.02)	.74	(.02)	.80	(.03)	93,975
B: Specification checks							
i) Fitted polynomial	-.35	(.02)	.72	(.01)	.78	(.03)	93,975
C: Work by end of year (2015-2016)	-.39	(.02)	.75	(.02)	.84	(.04)	68,288
i) Working next year	-.44	(.02)	.82	(.02)	.85	(.03)	50,791
a) In same firm	-.45	(.02)	.84	(.02)	.86	(.03)	46,162
b) In different firm	-.36	(.06)	.77	(.04)	.76	(.09)	10,573
ii) Not working next year	-.20	(.05)	.32	(.04)	.98	(.15)	17,497

Notes: This table displays estimates of labor force participation responses ($\triangle LFP$) and the corresponding participation elasticity wrt. the (net-of) participation tax rate (ϵ). $\triangle LFP$ is defined as the difference between the observed probability of working by the end of the calendar year (LFP) and the estimated counterfactual probability ($\hat{LFP}$) at the \$8,000 kink. The participation elasticity (ϵ) is the elasticity of labor force participation wrt. the (net-of) participation tax rate as defined in equation 3. In panel A, C and D, $\hat{LFP}$ is estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink. In Panel B i), $\hat{LFP}$ is estimated by a flexible polynomial as described in equation 2 using a 7th degree polynomial. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

4.1.2 Population Shares

Next, we relate our estimates of missing mass to the year-end excess mass of earnings around the kink. To obtain an estimate of the overall prevalence of hours frictions, we scale these two ratios by the frequency of each bin at the kink for employment and earnings.

Our results show clear signs of bunching at the kink by the end of the calendar year. Panel (b) in Figure 4 displays the observed year-end earnings distribution around the kink and the estimated counterfactual density. We estimate the counterfactual density using a seventh-degree polynomial, omitting observations in the excluded region. Using this specification, we estimate the excess mass to be 2.79, meaning that there are almost three times as many individuals with earnings at (or just around) the kink as predicted by the polynomial function. This estimate is bootstrapped 200 times, and the standard error confirms that it is highly significant.

Weighted by the frequency in the bin, the excess mass amounts to 3,151 individuals. By combining this with our estimate of the number of "missing workers" from Panel (a), we find that the estimated fraction of individuals who are hours constrained is 1,869/3,151 = 0.59, with a standard error of approximately 0.20. Our estimates indicate that a substantial fraction of individuals face hours constraints, suggesting that 59 percent of those who would optimally spread their hours more evenly throughout the calendar year are unable to do so and instead stop working due to hours frictions.

Table 3 reports our estimates of missing mass and the fraction of hours-constrained workers. Row A presents the baseline estimates shown in Figure 4, with our baseline estimate of the fraction of hours-constrained workers at 59 percent.

A potential concern is that some individuals may quit the labor market entirely, potentially biasing the estimate upward. To address this concern, we report estimates that include only individuals who continue working in the next calendar year in row B. Since our data span through 2017, we restrict this specification to observations from 2015–2016. Our results indicate that the estimated responses are higher for individuals who are observed to only temporarily leave the labor market. In fact, we estimate a higher fraction of hours-constrained workers at approximately 77 percent. This suggests that the responses are indeed driven by individuals who temporarily exit the labor market rather than those who leave permanently.

Table 3: Missing Mass and the Prevalence of Hours Constraints: Kink in the Disability Insurance Program ($8,000)

	Missing mass (m)		Year-end bunching (b_{12})		Fraction hours constrained (α)		Obs
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A: Baseline (2015-2017)	2.35	(.25)	2.79	(.13)	.59	(.12)	93,975
B: Working next year (2015-2016)	2.88	(.23)	2.93	(.17)	.77	(.12)	50,791
i) In same firm	2.91	(.22)	3.01	(.17)	.78	(.12)	46,162
ii) In different firm	3.00	(.62)	2.09	(.34)	1.02	(.44)	10,573

Notes: This table displays the estimated missing work, year-end bunching and fraction hours constrained for the sample of DI recipients. Missing work (m) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region around the kink relative to $\hat{LFP}$ at the kink, and is weighted by the estimated counterfactual cumulative earnings frequency. $\triangle LFP$ is defined as the difference between LFP and $\hat{LFP}$ at the $8,000 kink. Year-end bunching (b_{12}) is defined as the total difference between the observed and counterfactual year-end earnings distribution in the excluded region around the kink relative to the estimated counterfactual at the kink. Fraction hours constrained (α) is defined as the fraction of missing man-months (m_n) relative to the number of year-end bunchers ($b_{n,12}$). For the full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

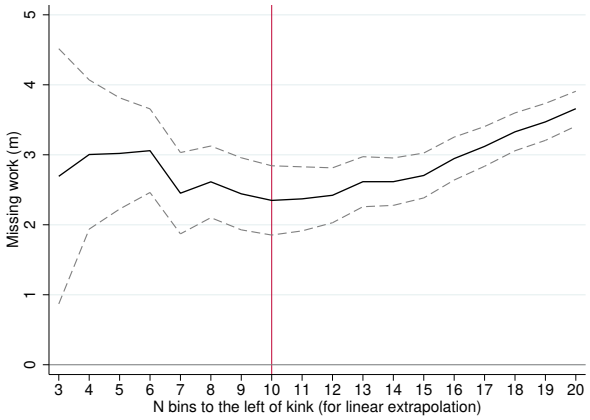
The estimates remain robust when restricting the sample to individuals who only temporarily stop working at the same firm, as shown in row B(i). Consistent with theory, we find that quitting work and starting at a new firm the following year is associated with more binding hours constraints. As shown in row B(ii), our estimate is effectively 100 percent, though we note that these estimates are considerably noisier. Overall, our evidence show that hours constraints are prevalent among DI recipients. The substantial missing mass in the earnings distribution, combined with high estimates of the fraction of hours-constrained workers, suggests that many individuals stop working mid-year in response to the strong benefit offset incentive.

4.1.3 Robustness

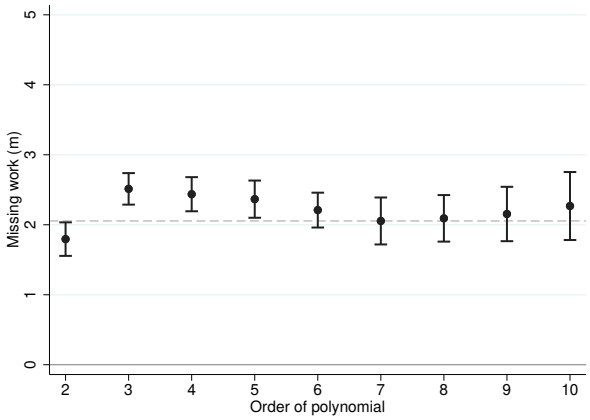
We show that our main findings are robust to alternative specifications. Figure 5 Panel (a) shows that our estimates are robust to the number of bins we use to the left of the excluded region using the linear extrapolation method, which changes the slope of the counterfactual probability and in turn changes the estimated responses. Panel (b) shows that our estimates remain quantitatively similar if we change the order of the polynomial we use to fit the counterfactual probability of continued work using the polynomial method.

Figure 5: Sensitivity to Counterfactual Specification in Missing Mass Estimation

(a) Linear Extrapolation: Sensitivity



(b) Polynomial Method: Sensitivity



Notes: Figure A displays the estimated missing work (m) for the sample of DI recipients where the counterfactual labor force participation ($\hat{LFP}$) is estimated by a linear extrapolation ($i = 1$) as described in equation 2 by using different numbers of bins to the left of the kink to estimate $\hat{LFP}$ (indicated by the horizontal axis). The red vertical line indicates our main specification using 10 bins to the left of the kink to estimate $\hat{LFP}$. Similarly, Figure B displays the difference between observed LFP and an estimated counterfactual $\hat{LFP}$ which is estimated by a n-th degree polynomial omitting bins in the excluded region as described in equation 2, where the degree of the polynomial (n) is indicated by the horizontal axis. The dashed horizontal line indicates our main specification using a 7-th degree polynomial to estimate $\hat{LFP}$. Both specifications use a 7-th degree polynomial to estimate the counterfactual cumulative earnings density (in order to estimate m). For full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (indicated by the dashed lines and vertical bars) are calculated by bootstrap using 500 replications.

Appendix Figure A.2 shows that our estimates are fairly robust when using a flexible polynomial (omitting observations in the excluded region) to estimate the counterfactual labor force participation rate, compared to the baseline linear extrapolation method using observations to the left of the excluded region. In Appendix Figure A.3 Panel (a), we show that the estimated year-end bunching is fairly robust to changing the order of the polynomial used to estimate the counterfactual year-end density, although estimates are somewhat higher for polynomials larger than 7. However, our estimate of the fraction of hours constrained workers is not qualitatively changed by choosing other orders of polynomials as shown in Panel (b), with point estimates ranging between about 0.45–0.75. Importantly, estimates are still highly significantly different from zero using other orders

of polynomials.

Finally, some may prefer to start working in February and continue working until December, rather than start working in January and stop working in December. Through the lens of our theory, this behavior can be rationalized as a knife-edge case of no-discounting. Alternatively, we need seasonal variation in search or quitting costs, for example, that it is easier to find a job starting in February than in January. It is also possible that some may prefer to stop and start multiple times during the year. While our approach is unable to capture this behavior, it does not invalidate our estimates of the participation response, but it may imply that we are underestimating the fraction hours constrained. To gauge this possibility, we implement an alternative estimation approach suggested by our theory. As indicated in Figure 1, the fraction constrained can also also be inferred by comparing the excess mass of cumulative earnings in the last two months of the year, as the bunching before the final period must be due to hours constraints. Reassuringly, comparing excess mass in November (1.56) to December (2.26), shown in A.5, yields a similar estimate of 0.69, albeit slightly larger.

4.1.4 Worker and Firm Heterogeneity

To assess the role of heterogeneity among our group of marginally attached workers, Table 4 reports estimates of the participation elasticity, missing mass, excess mass, and the implied fraction hours constrained for different subgroups. We find that the participation elasticity is substantially larger among the older subsample compared to the young. This may indicate that younger workers have a stronger incentive to return to the workforce as they have a potentially longer time until retirement.

Across the remaining subgroups, we find less stark differences. For example, there are no significant differences in responses between men and women, or between individuals with different educational attainment or types of work contracts. However, we find evidence that the participation response is greater in smaller firms that large firms, which may indicate that large firms have more career opportunities or flexibility.

Table 4: Heterogeneity in Missing Mass

	Participation elasticity (ϵ)		Missing Mass (m)		Year-end bunching (b_{12})		Fraction hours constrained (α)		Obs
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
All working DI recipients	.80	(.03)	2.35	(.25)	2.79	(.13)	.59	(.12)	93,975
Age 18-49	.57	(.05)	1.14	(.28)	1.08	(.19)	.76	(.40)	35,128
Age 50-67	.89	(.04)	3.00	(.31)	3.73	(.19)	.56	(.12)	58,847
Male	.85	(.04)	2.58	(.29)	2.72	(.16)	.64	(.14)	43,748
Female	.75	(.05)	2.12	(.34)	2.86	(.18)	.55	(.16)	50,227
No high school	.78	(.05)	2.21	(.26)	2.22	(.20)	.71	(.17)	48,771
High school or more	.83	(.04)	2.57	(.34)	3.41	(.18)	.53	(.13)	45,204
Regular pay	.79	(.05)	2.02	(.35)	3.01	(.26)	.59	(.19)	20,459
Some variable pay	.80	(.04)	2.75	(.33)	3.01	(.18)	.59	(.13)	60,666
Firm characteristics:									
% full-time work: > median	.87	(.04)	3.23	(.31)	3.52	(.21)	.67	(.14)	46,988
% full-time work: < median	.70	(.05)	1.43	(.31)	2.06	(.20)	.48	(.17)	46,987
% fixed pay: > median	.76	(.05)	2.40	(.30)	2.84	(.20)	.64	(.16)	48,236
% fixed pay: < median	.85	(.04)	2.36	(.30)	2.72	(.19)	.57	(.13)	45,739
% hourly pay: > median	.86	(.04)	2.65	(.34)	3.53	(.18)	.52	(.13)	47,011
% hourly pay: < median	.71	(.05)	1.99	(.31)	1.82	(.18)	.79	(.26)	46,964
N employees: 20 or less	.91	(.04)	3.20	(.43)	4.30	(.21)	.53	(.14)	34,649
N employees: 21-100	.71	(.06)	1.90	(.32)	1.31	(.22)	1.06	(.42)	33,791
N employees: More than 100	.66	(.07)	1.84	(.44)	2.08	(.25)	.63	(.27)	25,535

Notes: This table displays estimated missing work, year-end bunching, and the fraction hours constrained for different subgroups of DI recipients. For the full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

4.2 Responses to Kinks in the Income Tax Schedule

Having established that participation effects are large among marginally attached workers, we turn to an analysis of the income tax system. We present evidence across the full spectrum of kinks, covering the bottom to the top of the income distribution, with marginal tax changes ranging from 3 to 25 percentage points. We begin by plotting the participation probability function and the earnings distribution for each tax kink in Appendix Figure A.6, along with estimates of the missing mass and the participation effect.

At the lowest kink, at $6,667, we find evidence of some missing mass of work. Our estimate implies that about 15 percent stop working at the kink. As described in the previous section, this fraction is estimated by subtracting the height of the observed bins in the excluded region and the estimated counterfactual height. The counterfactual is estimated and fitted using the upward trending fraction who work, where we exclude the three bins of cumulative earnings closest to the kink. To contrast the missing mass of work, we also

plot the year-end cumulative distribution of earnings—including December—and an estimate of the excess mass of earnings around the kink. By contrast, we find no evidence of missing mass of work or excess mass of earnings at the next three kinks between $9,851 and $27,612.

At the fifth—or second highest—kink at $73,407, we see that the fraction who continue working is virtually flat on either side of the kink, indicating that the counterfactual can be estimated by simply fitting a line from the right or from the left. Comparing the drop at the kink to the fitted counterfactual, there is some evidence of missing mass – although the missing mass and participation effects are quite small – they are statistically significant different from zero. However, the noisy estimates of excess mass in year-end earnings prevents us from making a conclusions about the overall prevalence of hours constraints.

Table 5: Estimates From The Tax System

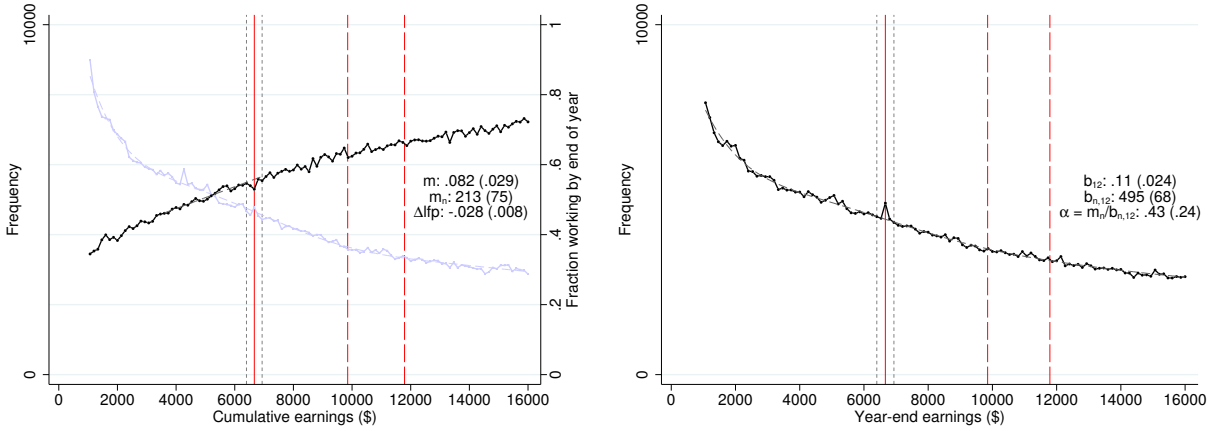
	mtr	$\triangle mtr$	Participation elasticity (ϵ)		Missing mass (m)		Year-end bunching (b_{12})	
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Kink point:								
$6,667	25	25	.20	(.06)	.082	(.029)	.115	(.024)
$9,851	8.2	-16.8	-.08	(.06)	.020	(.023)	-.059	(.024)
$11,789	23.6	15.4	.01	(.09)	.017	(.027)	.043	(.025)
$27,612	35.2	11.6	-.05	(.06)	-.001	(.019)	.009	(.028)
$73,407	44.2	9	.06	(.02)	.016	(.006)	.021	(.023)
$118,080	47.2	3	-.12	(.11)	-.008	(.012)	.142	(.055)
Placebo kinks:								
100,000 NOK	23.6	0	-.020	(.010)	.001	(.029)	-.010	(.025)
150,000 NOK	23.6	0	.016	(.007)	.000	(.024)	.087	(.027)
200,000 NOK	23.6	0	.010	(.008)	-.025	(.020)	.016	(.030)

Notes: This table displays the estimated participation elasticity, missing work, and year-end bunching for the workforce-sample (non-DI) for each kink in the income tax schedule. The participation elasticity (ϵ) is the elasticity of labor force participation wrt. the (net-of) participation tax rate as defined in equation 3. Missing work (m) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region around each kink point relative to $\hat{LFP}$ at the kink, and is weighted by the estimated counterfactual cumulative earnings frequency. Year-end bunching (b_{12}) is defined as the total difference between the observed and counterfactual year-end earnings distribution in the excluded region around each kink relative to the estimated counterfactual at the kink. For the full details of the estimated parameters, see section 2.2. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Figure 6: Graphical Representation of Hours Constraints: First Tax Kink ($6,667)

(a) Missing Mass

(b) Year-End Bunching (b_{12}) and Fraction Hours Constrained (α)



Notes: Figure A displays the observed cumulative earnings distribution for the workforce-sample (non-DI) and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). The figure further displays the observed fraction of recipients who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of $133 = 1,000 NOK) and the estimated counterfactual probability ($\hat{LFP}$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axis). Missing work (m) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region around the $6,667 kink relative to $\hat{LFP}$ at the kink, and is weighted by the estimated counterfactual frequency. The excluded region is indicated by the dashed vertical lines. Similarly, missing man-months (m_n) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region, weighted by the estimated counterfactual frequency. $\triangle LFP$ is defined as the difference between LFP and $\hat{LFP}$ at the $6,667 kink. Figure B displays the observed year-end earnings distribution and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). Year-end bunching (b_{12}) is defined as the total difference between the observed and counterfactual distribution in the excluded region around the $6,667 kink relative to the counterfactual at the kink. Similarly, the number of year-end bunchers ($b_{n,12}$) is defined as the total difference between the observed and counterfactual distribution in the excluded region. In both figures, the red dashed lines indicate (other) kinks in the income tax schedule at $9,851 and $11,789 respectively. For the full details of the estimated parameters, see section 2.2. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

5 Conclusion

In this paper, we have explored how hours mismatch affects labor supply responses to annual taxes. We started by providing a theoretical framework that showed that hours constraints can induce workers to stop working mid-year as their cumulative taxable income crept into a new tax bracket. We developed an approach to identify the intertemporal extensive margin response non-parametrically using data on monthly work decisions and cumulative income.

We applied this approach to annual tax kinks and monthly earnings data from the Norwegian tax and transfer system. We estimated a labor force participation elasticity of 0.8 and found that approximately 60% of individuals were hours constrained. There is suggestive evidence that these responses are larger among workers in small firms and among older workers – groups likely to face lower career concerns than younger workers.

By using our estimate of the share of hours-constrained individuals, we show that in the limiting case where all workers are constrained, the participation elasticity exceeds 1. In contrast, when using broader tax variation, we find much smaller elasticities, consistent with higher search costs or lower utility costs of overwork among the broader population.

These findings have implications for understanding the sources of fluctuations emphasized in the real business cycle (RBC) literature. Chari *et al.* (2007) and Shimer (2009) document a sizable gap, commonly referred to as the "labor wedge"—between the marginal rate of substitution between consumption and leisure and the marginal product of labor. This wedge must vary cyclically to explain the magnitude of employment fluctuations observed in the data, which are difficult to reconcile with standard estimates of labor supply elasticities. For instance, Card & Hyslop (2005) estimate an intertemporal elasticity of 0.38 using a benefit phase-out policy in Canada—far below the elasticity of 2 required by canonical RBC models (see, e.g., Cho & Cooley, 1994 and King & Rebelo, 1999). Our findings for marginally attached workers point to substantially higher elasticities, helping to reconcile this discrepancy. Moreover, they suggest that expanding the safety net, while maintaining strong work incentives, can amplify the responsiveness of labor supply at the extensive margin, consistent with a labor wedge that drives cyclical employment fluctuations.

References

ATKINSON, A., RAINWATER, L., & SMEEDING, T.M. 1995. Income Distributions in OECD countries: Evidence from the Luxembourg Income Study. *OECD Publications and Information Center*.

BHARGAVA, SAURABH, & MANOLI, DAYANAND. 2015. Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *The American Economic Review*, **105**(11), 3489–3529.

CARD, DAVID, & HYSLOP, DEAN R. 2005. Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers. *Econometrica*, **73**(6), 1723–1770.

CHARI, V. V., KEHOE, PATRICK J., & MCGRATTAN, ELLEN R. 2007. Business Cycle Accounting. *Econometrica*, **75**(3), 781–836.

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

CHETTY, RAJ, & SAEZ, EMMANUEL. 2013. Teaching the tax code: Earnings responses to an experiment with EITC recipients. *American Economic Journal: Applied Economics*, **5**(1), 1–31.

CHETTY, RAJ, LOONEY, ADAM, & KROFT, KORY. 2009. Salience and taxation: Theory and evidence. *American economic review*, **99**(4), 1145–77.

CHETTY, RAJ, FRIEDMAN, JOHN, OLSEN, TORE, & PISTAFERRI, LUIGI. 2011. Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics*, **126**, 749–804.

CHETTY, RAJ, GUREN, ADAM, MANOLI, DAY, & WEBER, ANDREA. 2013. Does indivisible labor explain the difference between micro and macro elasticities? A meta-analysis of extensive margin elasticities. *NBER macroeconomics Annual*, **27**(1), 1–56.

CHO, JANG-OK, & COOLEY, THOMAS F. 1994. Employment and Hours Over the Business Cycle. *Journal of Economic Dynamics and Control*, **18**(2), 411–432.

DIAMOND, PETER. 1980. Income taxation with fixed hours of work. *Journal of Public Economics*, **13**(1), 101–110.

FINKELSTEIN, AMY. 2009. E-ztax: Tax salience and tax rates. *The Quarterly Journal of Economics*, **124**(3), 969–1010.

GELBER, ALEXANDER M, JONES, DAMON, & SACKS, DANIEL W. 2019. Estimating Earnings Adjustment Frictions: Method and Evidence from the Earnings Test. *American Economic Journal: Applied Economics*.

GOLDIN, CLAUDIA. 2014. A grand gender convergence: Its last chapter. *American economic review*, **104**(4), 1091–1119.

HANSEN, GARY D. 1985. Indivisible labor and the business cycle. *Journal of monetary Economics*, **16**(3), 309–327.

HASTINGS, JUSTINE S, & WEINSTEIN, JEFFREY M. 2008. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, **123**(4), 1373–1414.

JENSEN, ROBERT. 2010. The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, **125**(2), 515–548.

JONES, DAMON. 2012. Inertia and overwithholding: explaining the prevalence of income tax refunds. *American Economic Journal: Economic Policy*, **4**(1), 158–185.

KEANE, MICHAEL P. 2011. Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, **49**(4), 961–1075.

KING, ROBERT G., & REBELO, SERGIO T. 1999. Resuscitating Real Business Cycles. *Pages 927–1007 of:* TAYLOR, JOHN B., & WOODFORD, MICHAEL (eds), *Handbook of Macroeconomics*, vol. 1. Amsterdam: North Holland: Elsevier.

KLEVEN, HENRIK, KREINER, CLAUS THUSTRUP, LARSEN, KRISTIAN, & SØGAARD, JAKOB EGHOLT. 2025. Micro vs macro labor supply elasticities: The role of dynamic returns to effort. *American economic review*.

KLEVEN, HENRIK J., & WASEEM, MAZHAR. 2013. Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics*, **128**, 669–723.

KLINE, PATRICK, & TARTARI, MELISSA. 2016. Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *American Economic Review*, **106**(4), 972–1014.

KOPCZUK, WOJCIECH, & MUNROE, DAVID. 2015. Mansion tax: the effect of transfer taxes on the residential real estate market. *American economic Journal: economic policy*, **7**(2), 214–57.

KOSTØL, ANDREAS RAVNDAL, & MOGSTAD, MAGNE. 2014. How Financial Incentives Induce Disability Insurance Recipients to Return to Work. *American Economic Review*, **104**(2), 624–55.

KRUEGER, ALAN B, & MEYER, BRUCE D. 2002. Labor supply effects of social insurance. *Handbook of public economics*, **4**, 2327–2392.

LACHOWSKA, MARTA, MAS, ALEXANDRE, SAGGIO, RAFFAELE, & WOODBURY, STEPHEN A. 2023. *Work hours mismatch*. Tech. rept. National Bureau of Economic Research.

LIEBMAN, JEFFREY B., & LUTTMER, ERZO F.P. 2012. Would People Behave Differently If They Better Understood Social Security? Evidence From a Field Experiment. *American Economic Journal: Economic Policy*, **7(1)**(17287), 275–99.

MOFFITT, ROBERT. 1992. Incentive Effects of the U.S. Welfare System: A Review. *Journal of Economic Literature*, **30**(1), 1–61.

ROGERSON, RICHARD. 1988. Indivisible labor, lotteries and equilibrium. *Journal of monetary Economics*, **21**(1), 3–16.

RUH, PHILIPPE, & STAUBLI, STEFAN. 2019. Financial incentives and earnings of disability insurance recipients: Evidence from a notch design. *American Economic Journal: Economic Policy*, **11**(2), 269–300.

SAEZ, EMMANUEL. 2002. Optimal income transfer programs: intensive versus extensive labor supply responses. *The Quarterly Journal of Economics*, **117**(3), 1039–1073.

SAEZ, EMMANUEL. 2003. The effect of marginal tax rates on income: a panel study of bracket creep. *Journal of Public Economics*, **87**(5-6), 1231–1258.

SAEZ, EMMANUEL. 2010. Do taxpayers bunch at kink points? *American economic Journal: economic policy*, **2**(3), 180–212.

SAEZ, EMMANUEL, SLEMROD, JOEL, & GIERTZ, SETH H. 2012. The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of economic literature*, **50**(1), 3–50.

SHIMER, ROBERT. 2009. Convergence in Macroeconomics: The Labor Wedge. *American Economic Journal: Macroeconomics*, **1**(1), 280–297.

WEATHERS, ROBERT R, & HEMMETER, JEFFREY. 2011. The impact of changing financial work incentives on the earnings of Social Security Disability Insurance (SSDI) beneficiaries. *Journal of Policy Analysis and Management*, **30**(4), 708–728.

ZARESANI, AREZOU. 2020. Adjustment Costs and Incentives to Work: Evidence from a Disability Insurance Program. *Journal of Public Economics*, **188**(104223).

A Appendix:Additional Results

A.1 Tables

Table A.1: Missing Mass and the Prevalence of Hours Constraints: Frequencies

	Frequency of missing mass (m_n)		Frequency of year-end bunchers ($b_{12,n}$)		Fraction hours constrained (α)		Obs
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A: Baseline	1,869	(239)	3,151	(80)	.59	(.12)	93,975
B. Conditional on work next year	1,583	(154)	2,043	(66)	.77	(.12)	50,791
In same firm	1,513	(145)	1,934	(60)	.78	(.12)	46,162
In different firm	314	(78)	309	(28)	1.02	(.44)	10,573

Notes: This table displays the estimated frequency of missing work (m_n), year-end bunching ($b_{12,n}$), and fraction hours constrained ($\alpha = m_n/b_{12,n}$). The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Table A.2: Missing Mass at Kinks in the Income Tax Schedule

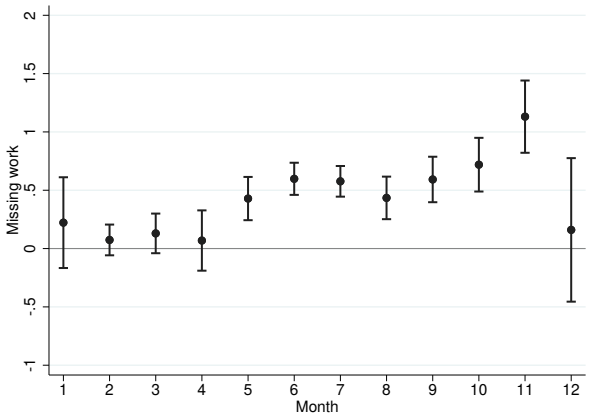
	mtr	$\triangle mtr$	Missing mass (m)		Year-end bunching (b_{12})		Obs
Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Kink point: 50,000 NOK ($6,667)	25	25	.136	(.028)	.21	(.02)	2,803,147
w/o round numbers (in multiples of 100 NOK)	25	25	.086	(.031)	.11	(.02)	2,803,147
w/o round numbers (in multiples of 1,000 NOK)	25	25	.082	(.029)	.11	(.02)	2,803,147
w/o round numbers (in multiples of 5,000 NOK)	25	25	.086	(.028)	.12	(.02)	2,803,147
w/o round numbers (in multiples of 10,000 NOK)	25	25	.075	(.030)	.13	(.02)	2,803,147
Placebo kink: 100,000 NOK	23.6	0	.051	(.026)	.08	(.03)	2,803,147
w/o round numbers (in multiples of 100 NOK)	23.6	0	.001	(.030)	-.01	(.02)	2,803,147
w/o round numbers (in multiples of 1,000 NOK)	23.6	0	.001	(.029)	-.01	(.03)	2,803,147
w/o round numbers (in multiples of 5,000 NOK)	23.6	0	-.001	(.026)	-.02	(.03)	2,803,147
w/o round numbers (in multiples of 10,000 NOK)	23.6	0	-.005	(.025)	-.01	(.02)	2,803,147
Placebo kink: 150,000 NOK	23.6	0	.034	(.023)	.17	(.03)	2,803,147
w/o round numbers (in multiples of 100 NOK)	23.6	0	-.005	(.025)	.07	(.03)	2,803,147
w/o round numbers (in multiples of 1,000 NOK)	23.6	0	.000	(.024)	.09	(.03)	2,803,147
w/o round numbers (in multiples of 5,000 NOK)	23.6	0	-.005	(.024)	.08	(.03)	2,803,147
w/o round numbers (in multiples of 10,000 NOK)	23.6	0	-.006	(.026)	.08	(.03)	2,803,147
Placebo kink: 200,000 NOK	23.6	0	.009	(.023)	.11	(.03)	2,803,147
w/o round numbers (in multiples of 100 NOK)	23.6	0	-.025	(.022)	.01	(.03)	2,803,147
w/o round numbers (in multiples of 1,000 NOK)	23.6	0	-.025	(.020)	.02	(.03)	2,803,147
w/o round numbers (in multiples of 5,000 NOK)	23.6	0	-.025	(.021)	.02	(.03)	2,803,147
w/o round numbers (in multiples of 10,000 NOK)	23.6	0	-.028	(.021)	.03	(.03)	2,803,147

Notes: This table displays the estimated missing work (m) and year-end bunching (b_{12}) for different sample restrictions in round numbers for the workforce-sample (non-DI). The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

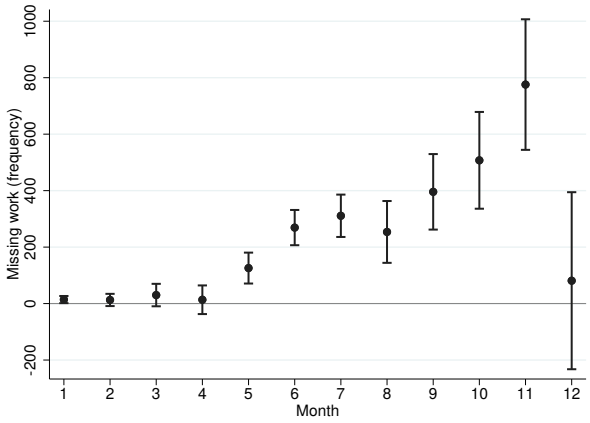
A.2 Figures

Figure A.1: Dynamics in Missing Mass: Benefit Phase-Out Kink in the DI Program (at $8,000)

(a) Missing Mass (m)



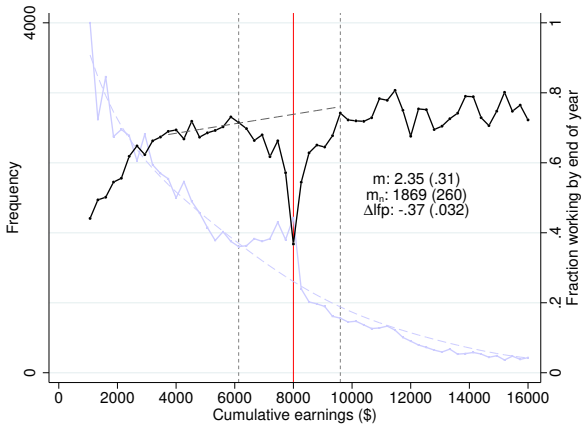
(b) Missing Mass: Frequency (m_n)



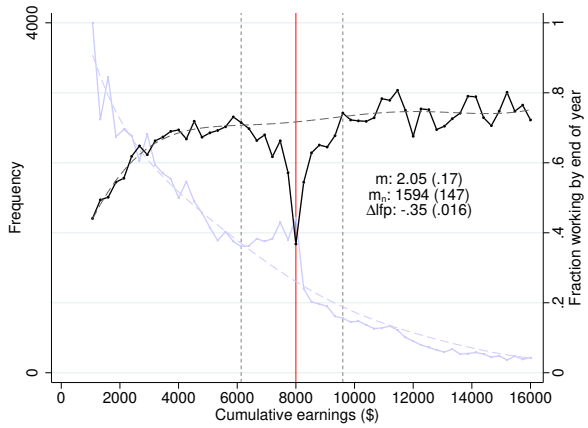
Notes: Figure A displays the estimated missing work (m) for the sample of DI recipients for each calendar month. m is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region around the kink relative to $\hat{LFP}$ at the kink, and is weighted by the estimated counterfactual frequency. Similarly, Figure B displays the estimated missing man-months (m_n) for each calendar month. m_n is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region, weighted by the estimated counterfactual frequency. For the full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (indicated by the vertical bars) are calculated by bootstrap using 500 replications.

Figure A.2: Missing Mass Estimation Methods

(a) Linear Extrapolation



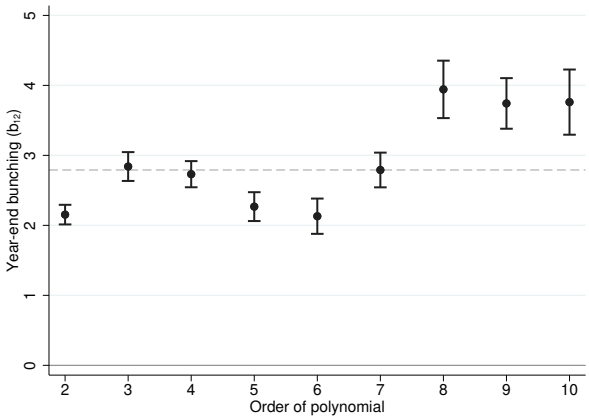
(b) Polynomial Method



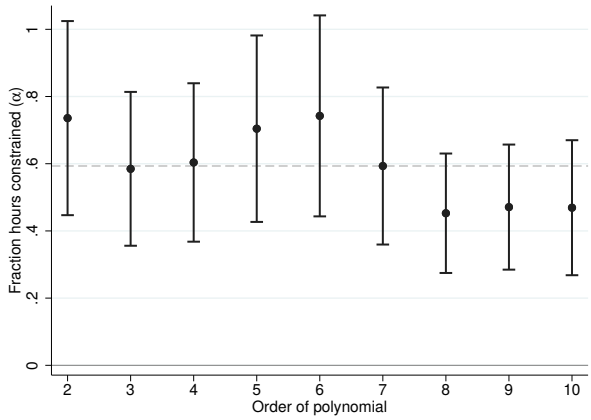
Notes: Figure A displays the observed fraction of recipients who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$267 = 2,000 NOK) and the counterfactual probability ($L\hat{F}P$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axis). Similarly, Figure B displays the observed LFP and estimated counterfactual $L\hat{F}P$ which is estimated by a 7th degree polynomial omitting bins in the excluded region as described in equation 2. In both figures, the figures display estimated parameters using each respective estimation method and the observed cumulative earnings distribution for the sample of DI recipients and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). For full details of the estimated parameters, see section 2.2. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Figure A.3: Sensitivity in Order of Polynomial in Bunching Estimation

(a) Year-End Bunchers (b_{12})



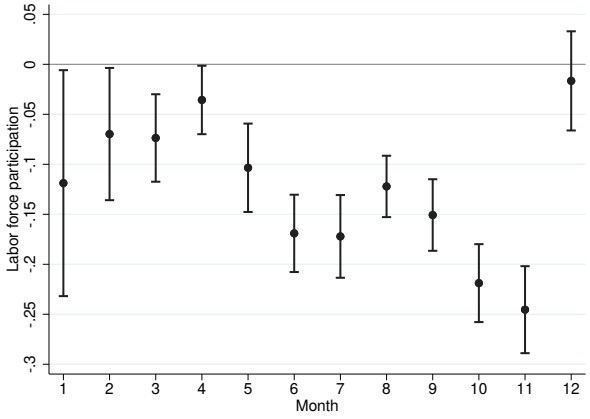
(b) Fraction Hours Constrained (α)



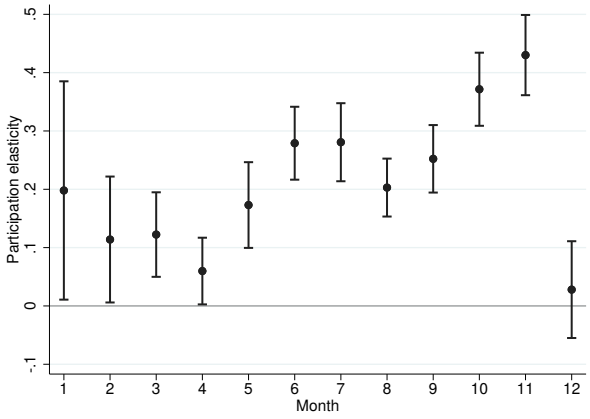
Notes: Figure A displays estimated year-end bunching (b_{12}) for the sample of DI recipients where the counterfactual earnings distribution is estimated by a n-th degree polynomial (indicated on the horizontal axis). Similarly, Figure B displays the estimated fraction hours constrained (α) for different orders of polynomial to estimate the number of year-end bunchers. For the full details of the estimated parameters, see section 2.2. In both figures, the dashed horizontal lines indicate our main specification using a 7-th degree polynomial to estimate the counterfactual density. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (indicated by the vertical bars) are calculated by bootstrap using 500 replications.

Figure A.4: Dynamics in Missing Mass: Benefit Phase-Out Kink in the DI Program (at $8,000)

(a) Labor Force Participation (ΔLFP)



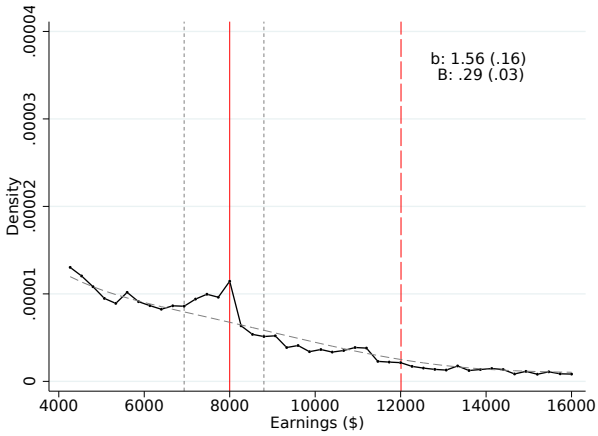
(b) Participation Elasticity (ϵ)



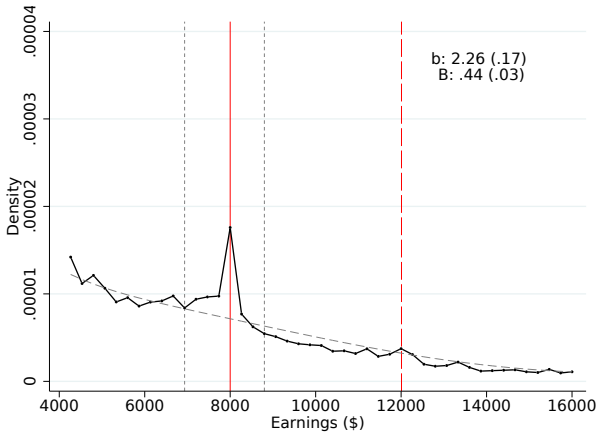
Notes: Figure A displays the estimated labor force participation responses ($\triangle LFP$) for the sample of DI recipients for each calendar month.$\triangle LFP$ is defined as the difference between the observed probability of working by the end of the calendar year (LFP) and the estimated counterfactual probability ($\hat{LFP}$) at the $8,000 kink. $\hat{LFP}$ is estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2. Figure B displays the estimated participation elasticity (ϵ) for each calendar month defined as the elasticity of LFP wrt. the (net-of) participation tax rate as defined in equation 3. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (indicated by the vertical bars) are calculated by bootstrap using 500 replications.

Figure A.5: Smooth vs Lumpy Bunching

(a) November



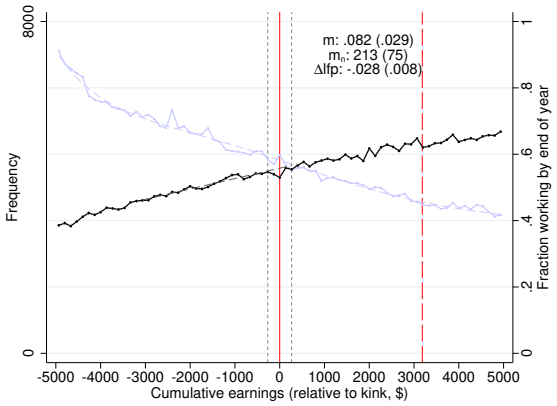
(b) December



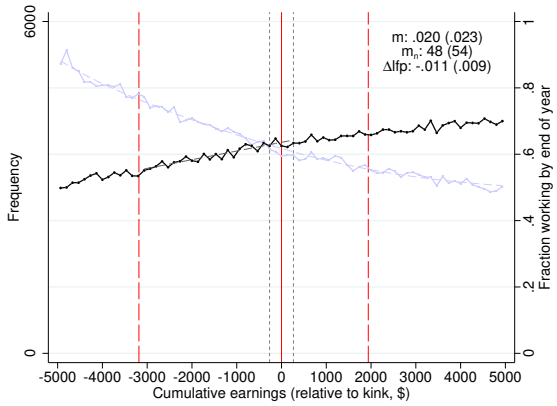
Notes: These figures display the observed (cumulative) earnings distributions in November and December, respectively, and the estimated bunching. The sample consists of DI recipients aged 18-66 who receive full (100 percent) benefits, have some labor market earnings during the calendar year, and who were awarded DI before 2015. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Figure A.6: Missing Mass at Kinks in the Income Tax Schedule

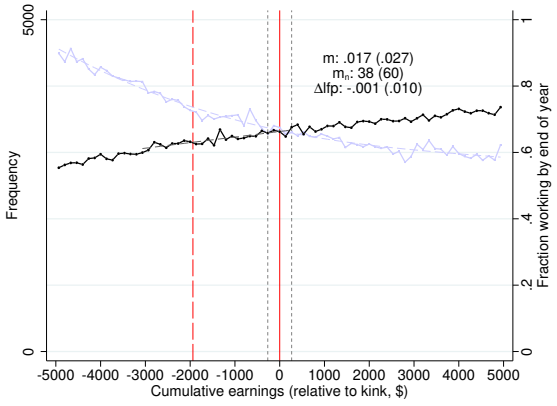
(a) $6,667 Kink



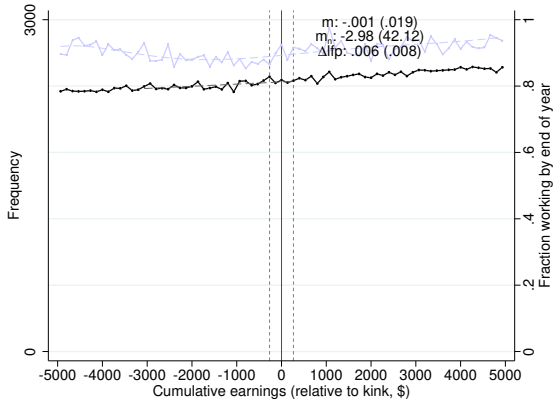
(b) $9,851 Kink



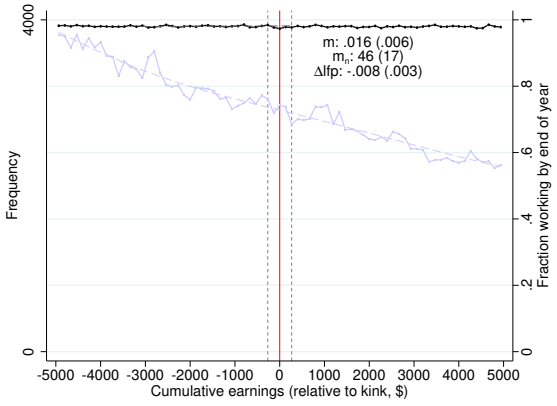
(c) $11,789 Kink



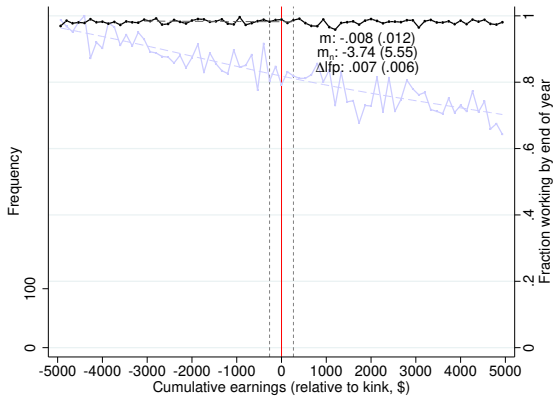
(d) $27,612 Kink



(e) $73,407 Kink

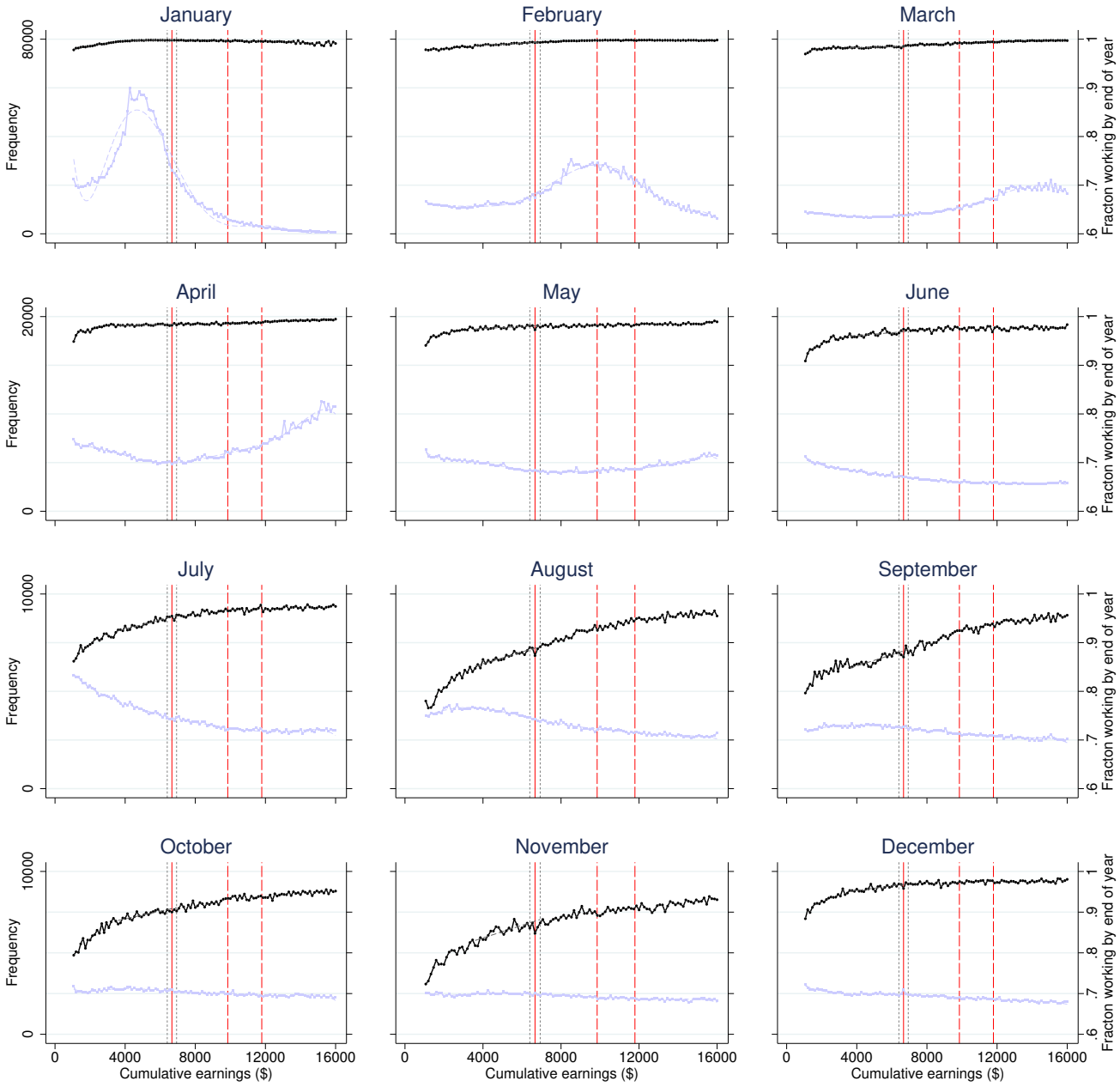


(f) $118,080 Kink



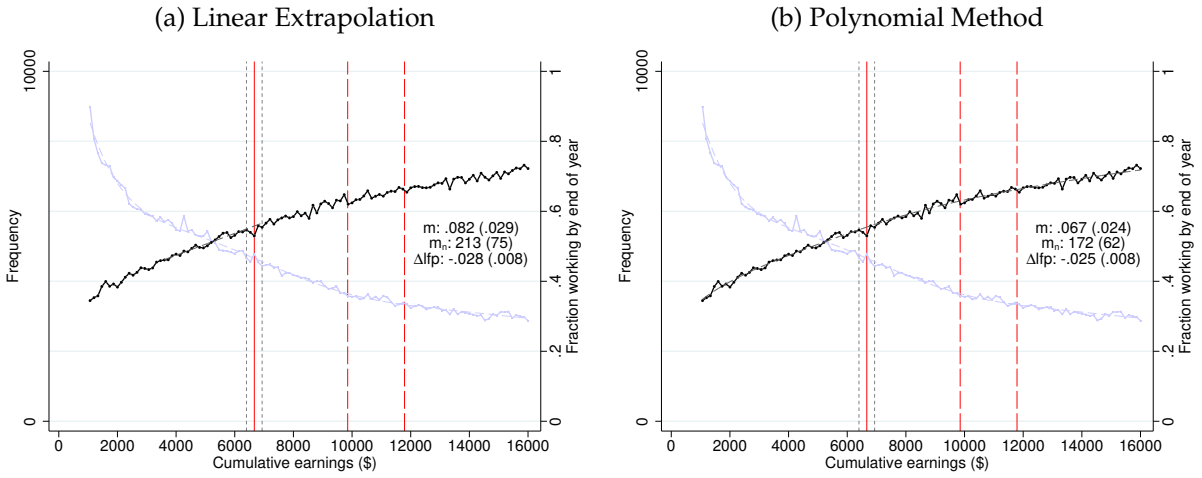
Notes: These figures display the observed cumulative earnings distributions around each kink in the income tax schedule for the workforce-sample (non-DI) along with fitted counterfactuals using a 7th degree polynomial as described in equation 2 (left axes). The figures further display the observed fraction of individuals who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$133 = 1,000 NOK) and the estimated counterfactual probabilities ($L\hat{F}P$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink. See Figure 4 for more details. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Figure A.7: Missing Mass by Month: First Kink in the Income Tax Schedule ($6,667)



Notes: These figures display the observed cumulative earnings distributions for the workforce-sample (non-DI) and fitted counterfactual distributions using a 7th degree polynomial as described in equation 2 for each calendar month (left axes). The figures further display the observed fractions of recipients who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$133 = 1,000 NOK) and the estimated counterfactual probabilities ($\hat{LFP}$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axes). Missing work (m) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region around the \$6,667 kink relative to $\hat{LFP}$ at the kink, and is weighted by the estimated counterfactual frequency. The excluded region is indicated by the dashed vertical lines. Similarly, missing man-months (m_n) is defined as the total weighted difference between LFP and $\hat{LFP}$ in the excluded region, weighted by the estimated counterfactual frequency. $\triangle LFP$ is defined as the difference between LFP and $\hat{LFP}$ at the \$6,667 kink. For the full details of the estimated parameters, see section 2.2. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

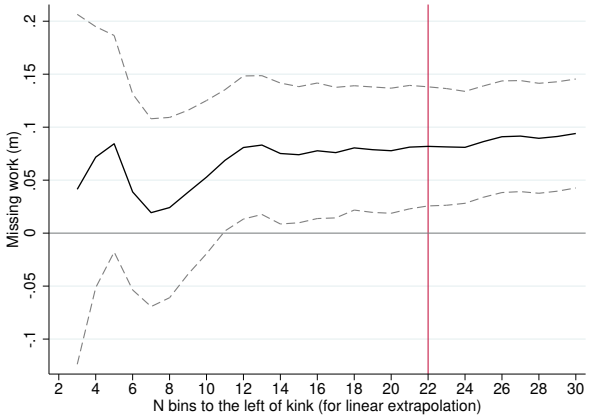
Figure A.8: Missing Mass Estimation Methods: First Kink in the Income Tax Schedule ($6,667)



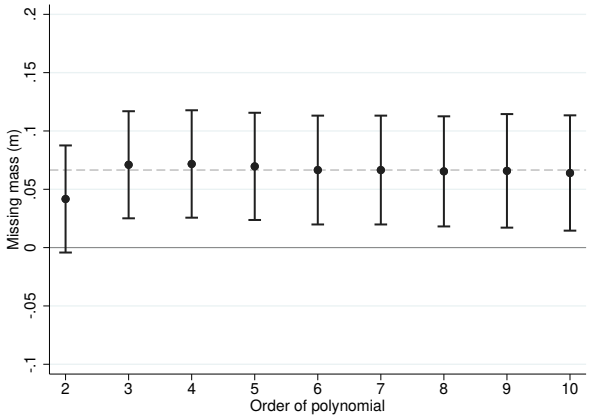
Notes: Figure A displays the observed fraction of individuals who work by the end of the calendar year (LFP) for each earnings bin (using a binwidth of \$133 = 1,000 NOK) and the counterfactual probability ($\hat{LFP}$) estimated by a linear extrapolation ($i = 1$) using observations to the left of the kink as described in equation 2 (right axis). Similarly, Figure B displays the observed LFP and estimated counterfactual $\hat{LFP}$ which is estimated by a 7th degree polynomial omitting bins in the excluded region as described in equation 2. Both figures display estimated parameters using each respective estimation method and the observed cumulative earnings distribution for the workforce-sample and a fitted counterfactual using a 7th degree polynomial as described in equation 2 (left axis). For full details of the estimated parameters, see section 2.2. In both figures, the red dashed lines indicate (other) kinks in the income tax schedule at \$9,851 and \$11,789 respectively. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (in parentheses) are calculated by bootstrap using 500 replications.

Figure A.9: Sensitivity to Counterfactual Specification in Missing Mass Estimation: First Kink in the Income Tax Schedule ($6,667)

(a) Linear Extrapolation: Sensitivity



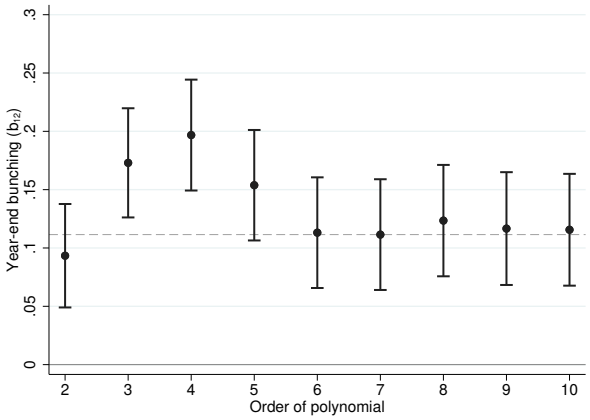
(b) Polynomial Method: Sensitivity



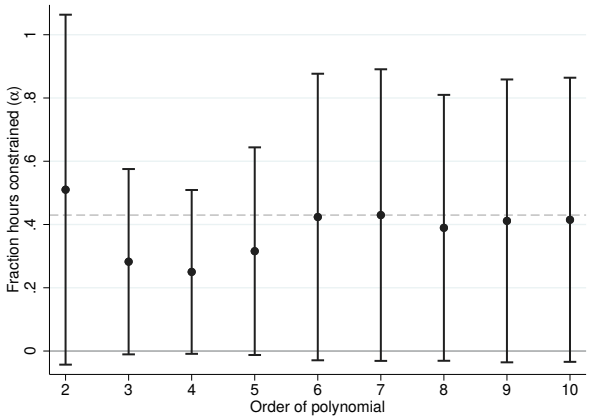
Notes: Figure A displays the estimated missing work (m) for the workforce-sample (non-DI) where the counterfactual labor force participation ($L\hat{F}P$) is estimated by a linear extrapolation by using different number of bins to the left of the kink to estimate $L\hat{F}P$ (indicated by the horizontal axis). The red vertical line indicates our main specification using 22 bins to the left of the kink to estimate $L\hat{F}P$. Similarly, Figure B displays the observed LFP and estimated counterfactual $L\hat{F}P$ which is estimated by a n-th degree polynomial omitting bins in the excluded region as described in equation 2, where the degree of the polynomial (n) is indicated by the horizontal axis. The dashed horizontal line indicates our main specification using a 7-th degree polynomial to estimate $L\hat{F}P$. Both specifications use a 7-th degree polynomial to estimate the counterfactual cumulative earnings density (in order to estimate m). For full details of the estimated parameters, see section 2.2. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (indicated by the dashed lines and vertical bars) are calculated by bootstrap using 500 replications.

Figure A.10: Sensitivity to Order of Polynomial in Bunching Estimation: First Kink in the Income Tax Schedule ($6,667)

(a) Year-End Bunchers ($b_{n,12}$)



(b) Fraction Hours Constrained (α)



Notes: Figure A displays estimated year-end bunching (b_{12}) for the workforce-sample (non-DI) where the counterfactual earnings distribution is estimated by a n-th degree polynomial (indicated on the horizontal axis). Similarly, Figure B displays the estimated fraction hours constrained (α) for different orders of polynomial to estimate the number of year-end bunchers. For the full details of the estimated parameters, see section 2.2. In both figures, the dashed horizontal lines indicate our main specification using a 7-th degree polynomial to estimate the counterfactual density. The sample consists of the overall workforce (non-DI) with some labor market earnings during the calendar year during 2015-2017. Standard errors (indicated by the vertical bars) are calculated by bootstrap using 500 replications.

B Theory

Proposition 1 follows from two lemmas.

Lemma 1. *If her total earnings hit strictly below the kink, it is optimal for the flexible worker to work every month.*

Proof. If total earnings hit strictly below the kink, a necessary condition at the optimum is

$$\delta^{t-1}\left[wu'\left(h_t w + x\right) - \theta D'\left(h_t\right)\right] = 0,$$

which is time invariant (as we can just divide out by the discount factor). Accordingly, if total earnings hit strictly below the kink, it is optimal for the worker to work all year. □

Lemma 2. *If a flexible worker's total earnings hit above the kink, it is optimal to stop mid-year only if her marginal disutility of work at* 0, $D'(0)$, *is strictly positive .*

Proof. There are two cases. First, let a flexible worker's total earnings hit the kink exactly, i.e., $w\sum_{i=1}^{12} h_i = K$. We will show that $h_t > 0$ for all t. Suppose for the sake of contradiction not: suppose first there exists some t and $t' > t$ such that $h_t = 0$ but $h_{t'} = h > 0$. We perturb the worker's payoff as follows. The derivative of

$$u\left(h_t w + x\right) - \theta D\left(h_t\right) + \delta^{t'-t}\left[u\left((h - h_t)w + x\right) - \theta D\left(h - h_t\right)\right],$$

with respect to h_t, evaluated at $h_t = 0$, is

$$wu'\left(x\right) - \theta D'\left(0\right) - \delta^{t'-t}\left[wu'\left(hw + x\right) - \theta D'\left(h\right)\right] > 0,$$

provided $D'(0)$ is sufficiently low, contradicting the optimality of $h_t = 0$. Second, suppose there exists some t' such that $h_t = 0$ for all $t > t'$. Pick an arbitrary $t > t'$. The derivative of

$$u\left(h_t w\left(1 - \tau\right) + x\right) - \theta D\left(h_t\right),$$

with respect to h_t, evaluated at $h_t = 0$, is

$$w(1 - \tau)u'\left(x\right) - \theta D\left(0\right) > 0,$$

provided $D'(0)$ is sufficiently low, contradicting the optimality of $h_t = 0$.

The second case is that in which a flexible worker's total earnings strictly exceed the

kink. Let s ($s \in \{0, \ldots, 11\}$) denote the last period before earnings strictly exceed the kink:

$$s := \max \left\{ t \in \{0, \ldots, 11\} \ : \ w \sum_{i=1}^{t} h_i \leq K \right\}.$$

By an identical perturbation argument as above, and by the definition of s, it must be the case that$h_t > 0$ for all $t \leq s + 1$. But then the derivative of the objective with respect to h_t for any $t \geq s + 2$, evaluated at $h_t = 0$, is

$$\delta^{t-1} \left(w \left(1 - \tau \right) u' \left(x \right) - \theta D' \left(0 \right) \right),$$

which is weakly less than 0 if and only if

$$\theta D' \left(0 \right) \geq w \left(1 - \tau \right) u' \left(x \right).$$

As the right hand side of this inequality is strictly positive, the result follows. □

B.1 Benefits Reduction

In the text, we remarked that there is little qualitative difference between a tax and a benefits reduction. We now elaborate on this comment. We assume that the tax is implemented via reduction in future benefits as follows. Let the worker be eligible for a nominal annual benefit of $12x$, and let

- $y_t = \sum_{i=1}^{t} h_i w$ denote cumulative labor income through period t,
- x_t denote the benefit paid in period t,
- X_t denote cumulative benefits paid through t,
- K denote the maximum cumulative labor income before benefit reduction begins, and
- $\tau \in (0, 1)$ denote the clawback rate applied to income above K.

Define the allowable annual income at time t , A_t:

$$A_t = \begin{cases} y_t + 12x, & \text{if} \quad y_t \leq K \\ K + 12x + (y_t - K)(1 - \tau), & \text{if} \quad K < y_t \end{cases}.$$

For all $j \geq t$, given cumulative benefits paid up to that point,X_{t-1}, the per-period benefit x_j is

$$x_j = \frac{A_t - y_t - X_{t-1}}{12 - t + 1}.$$

Observe that for all t such that $y_t \leq K$, $x_t = x$. Moreover, it is easy to see that when viewing the tax and benefits schedule from an annual perspective, this setup is identical to that in the text. However, there are subtle differences in our model with month-by-month labor supply decisions. In particular, the direct tax takes place "instantaneously," whereas the benefit reduction is spread out over the subsequent months. This, of course, only makes it more attractive for a flexible worker to work every period as this spreading diminishes the effective penalty to working (due to the concavity of the utility function and impatience).

Evidently, for a flexible worker, Lemma 1 still holds exactly (the tax- and benefits-reduction specifications are identical below the kink). There is also an analog of Lemma 2 :

Lemma 3. *If a flexible worker's total earnings hit above the kink and her marginal disutility of work at* 0 *is* 0 *(or sufficiently low), it is never optimal for her to stop mid-year.*

Proof. If earnings hit the kink exactly, i.e., $w\sum_{i=1}^{12} h_i = K$, we can mimic the first part of the proof of Lemma 2 nearly exactly to yield the result. Suppose now that it is optimal for earnings to strictly exceed the kink. Take an arbitrary period $t \geq s+1$ and let dU/dh_t denote the derivative of continuation utility from period t onward with respect to h_t, evaluated at $h_t = 0$, under the assumption that $D'(0) = 0$. Then,

$$\left.\frac{dU}{dh_t}\right|_{h_t=0} = \delta^{t-1} w u'(x_t)\left(1 - \frac{\tau}{12 - t + 1}\right) + \sum_{i=t+1}^{12} \delta^{i-1} u'(h_i w + x_i) \cdot \frac{dx_i}{dh_t}.$$

Observe that the first term is strictly positive, whereas each term in the sum is strictly negative. We now bound the magnitude of these spillover terms. As all $y_i > K$,

$$\left|\frac{dx_i}{dh_t}\right| \leq \frac{\tau w}{12 - t + 1} \quad \text{for all } i \geq t.$$

Moreover, by the concavity of u, we have $u'(h_i w + x_i) \leq u'(x_t)$. Thus,

$$\left|\sum_{i=t+1}^{12} \delta^{i-1} u'(h_i w + x_i) \cdot \frac{dx_i}{dh_t}\right| \leq u'(x_t) \cdot \frac{\tau w}{12 - t + 1} \cdot \sum_{i=t+1}^{12} \delta^{i-1}.$$

Let us compare this upper bound to the first term. Dividing out, it is easy to see that it

suffices to assume that $\delta = 1$. Accordingly, we have

$$1 - \frac{\tau}{12-t+1} > \frac{\tau(12-t)}{12-t+1} \quad \Leftrightarrow \quad 12-t+1-\tau > \tau(12-t) \quad \Leftrightarrow \quad (12-t+1)(1-\tau) > 0,$$

which holds for all $\tau \in (0,1)$. □

B.2 Relaxing Additive Separability

In the text, we specify the following period t utility

$$u\left(h_t w\left(1-\tau\left(y_t\right)\right)+x\right)-\theta D\left(h_t\right).$$

Let us now check that our result that flexible workers do not stop mid-year is not specific to this functional form. Indeed, suppose a worker's period t utility is instead

$$u\left(c_t, h_t; \theta\right),$$

which we assume to be

- continuously differentiable in both arguments;
- strictly increasing in consumption and strictly decreasing in labor: $\frac{\partial u}{\partial c}(c,h;\theta) > 0$ and $\frac{\partial u}{\partial h}(c,h;\theta) \leq 0$; and
- concave.

Then, letting

$$U_t = u\left(h_t w\left(1-\tau\left(y_t\right)\right)+x, h_t; \theta\right),$$

we have an analog of Lemmas 3 and 4:

Proposition 2. *For a flexible worker, if a worker's marginal disutility of work at* 0 *is* 0 *for all* $c \leq x$*, it is not optimal for her to stop mid-year.*

Proof. Suppose earnings hit above the kink and consider a period beyond the kink . Then the workers marginal utility at $h_t = 0$ is

$$\left.\frac{dU_t}{dh_t}\right|_{h_t=0} = w(1-\tau)\cdot\frac{\partial u}{\partial c}(x,0;\theta) + \frac{\partial u}{\partial h}(x,0;\theta).$$

We omit the arguments for the other periods (crossing and pre-crossing), as the logic is identical. □