

# Treatment Effects in Bunching Designs: The Impact of the Federal Overtime Rule on Hours

Leonard Goff\*

This version: December 28, 2021

[For current version click here](#)

## Abstract

The Fair Labor Standards Act (FLSA) mandates overtime premium pay for most U.S. workers, but a lack of changes to the rule has left limited opportunities to assess its impacts on the labor market. This paper uses bunching at 40 hours to estimate the effect of the FLSA overtime rule on hours of work, leveraging an extension of the “bunching design” identification strategy and a large sample of weekly paychecks. I develop a nonparametric approach to using bunching at a choice-set kink, partially identifying the average causal effect of the kink among bunchers given weak assumptions on choice. The bounds are informative in the overtime context and suggest that directly affected hourly workers in the U.S. work an average of at least half an hour less per week as a result of the FLSA mandate. This also delivers new estimates of the wage elasticity of hours demand.

---

\*Department of Economics, University of Georgia. Email: [leonard.goff@columbia.edu](mailto:leonard.goff@columbia.edu). I thank my co-advisors Simon Lee and Suresh Naidu, as well as Michael Best, Daniel Hamermesh, and Bernard Salanié for their for their gracious advice and support throughout this project. I have also benefited from discussions with Christopher Ackerman, Doug Almond, Joshua Angrist, Jushan Bai, Iain Bamford, Marc Bellemare, Sandra Black, Ivan Canay, Gregory Cox, Junlong Feng, Bhargav Gopal, Jonas Hjort, Wojciech Kopczuk, Bentley MacLeod, Matthew Masten, Serena Ng, José Luis Montiel Olea, Dilip Ravindran, Miguel Urquiola, and seminar participants at Columbia, Duke, Lehigh, Microsoft, the University of Georgia, the University of North Carolina Greensboro, the US Census Bureau, and Washington State University, as well as audiences at the Canadian Economics Association and Southern Economics Association meetings, and the European Winter Meeting of the Econometric Society. I thank my main data provider as well as the Bureau of Labor Statistics. Supplemental Material is available [here](#).

# 1 Introduction

Many countries require premium pay for long work hours, in an effort to limit excessive work schedules and encourage hours to be spread over more workers. In the U.S., such regulation comes through the “time-and-a-half” rule of the Fair Labor Standards Act (FLSA): workers must be paid one and a half times their normal hourly wage for any hours they work in excess of 40 within a single week. Although some workers are exempt from the overtime rule (and additionally some employers are not covered by the FLSA), the time-and-a-half rule applies to a majority of the U.S. workforce, including nearly all of its over 80 million hourly workers (U.S. Department of Labor, 2019). Workers in many industries average multiple overtime hours per week, making overtime the largest form of supplemental pay in the U.S. (Bishow, 2009).<sup>1</sup>

Nevertheless, only a small literature has addressed the causal effects of the FLSA overtime rule on the U.S. labor market. This stands in marked contrast to the large body of work on the minimum wage, which was also introduced at the federal level by the FLSA in 1938. A key reason for this gap is that the overtime rule has varied little since then: the main parameters have remained as time-and-a-half after 40 hours within a week, for now more than 80 years.<sup>2</sup> This lack of variation has afforded few opportunities to leverage research designs that exploit policy changes to identify causal effects.<sup>3</sup> In contrast to the minimum wage, reforms to overtime policy have been rare and have left the central parameters of the rule unaffected.

This paper takes a new approach to assessing the effect of the FLSA overtime rule on hours of work, by making use of variation *within* the rule itself: given a fixed hourly wage, hours in excess of 40 in a week from a single worker are more expensive to the firm than the first 40. This introduces a sharp discontinuity in the marginal cost of a worker-hour—a convex “kink” in firms’ total costs—which provides firms with an incentive to set workers’ hours exactly at 40. Models of optimizing behavior predict that the mass of workers with hours at 40 in a given week will be larger or smaller depending on how responsive firms are to the wage variation imposed by the FLSA. Combining this observation with assumptions about the shape of the distribution of hours that would be chosen in the absence of the FLSA, I use the bunching mass to identify the effect of the overtime rule on hours. This builds on a literature that has used bunching at kinks in tax liability to identify the

---

<sup>1</sup>Hart (2004) reports an average of 3 overtime hours per week among non-supervisory production workers. See Table F.2 for new estimates by industry from my sample. From a separate representative survey I estimate in Section 3 a grand average of about one overtime hour per week per worker, among all employed.

<sup>2</sup>While there are supplemental state overtime rules that vary somewhat by state (e.g. Minnesota has a 48 hour threshold), these rules bind for relatively few workers since the federal rules supersede the state rules.

<sup>3</sup>A notable exception is Hamermesh and Trejo (2003), who apply a difference-in-differences approach over the expansion of a daily overtime rule in California to men in 1980, estimating a price elasticity of demand for overtime hours of roughly  $-0.5$ . Costa (2000) and Johnson (2003) also consider the impact of the federal rule on hours, studying the initial phase-in of the FLSA and a supreme court decision clarifying the eligibility of public sector workers, respectively. Quach (2021) looks at very recent reforms to eligibility criteria for exemption from the FLSA, estimating effects on employment and incomes of salaried workers, but not on hours.

elasticity of labor supply (e.g. Saez 2010; Chetty et al. 2011),<sup>4</sup> which I refer to as the “bunching design”.<sup>5</sup>

In order to operationalize a bunching-design approach to assessing the overtime rule, this paper requires two new ingredients relative to existing literature. First, my approach relies on high-resolution data on the hours of individual workers within a single week, in order to observe the precise distribution of hours close to 40. I obtain this through a novel dataset of individual paycheck records from a large payroll processing company, which reports the exact number of hours that a worker was paid for in a given week. Second, my approach requires a way to translate this observed hours distribution into causal estimates of the rule’s effect, under credible assumptions about how weekly working hours are determined. This requires moving beyond the original bunching-design model popularized in public finance applications, which assumes a stylized labor supply model with isoelastic preferences and strong restrictions on heterogeneity.

The identifying assumptions of the bunching design can be separated into two parts: i) assumptions about how individual agents would make choices given counterfactual choice sets—a *choice model*, and ii) assumptions about the distribution of heterogeneity in choices across agents. I find that the class of choice models under which the bunching-design can be applied is considerably more general than has been previously appreciated. In particular, I show that the method does not require the researcher to suppose any explicit functional form for decision-makers’ utility; rather, the core behavioral prediction driving identification rests on *convexity* of individual preferences (and of the kink itself). In my formulation, agents in the bunching design may even have multiple underlying margins of choice, which can be unobserved to the researcher and vary by agent. These findings establish an important robustness property for the bunching design: it remains broadly valid even when the isoelastic utility model typically used to motivate the design is misspecified.

This generality is accomplished by recasting the bunching design in a potential outcomes framework, defining the parameter of interest in terms of counterfactual *choices* rather than as a preference parameter from a parametric choice model. I show that choice from a kinked choice set can be fully characterized by two such counterfactuals, and that bunching directly identifies a feature of their joint distribution. In the overtime setting I take firms to choose the hours of workers, and these potential outcomes correspond to: a) the number of hours the firm would choose for the worker if the worker’s normal wage rate applied to all hours this week; and b) the number that the firm would choose if the worker’s overtime rate applied to all hours this week.

In addition to generalizing the choice model underlying the bunching design, I show that the

---

<sup>4</sup>Variations of the bunching design have since been applied in a variety of settings from cell phone plan pricing (Huang, 2008) to fuel economy standards (Ito and Sallee, 2017), prescription drug spending (Einav et al., 2017) and Social Security (Gelber et al., 2020), among others.

<sup>5</sup>This paper considers only the bunching design for kinks, and not the related method for bunching at *notches* (e.g. Kleven and Waseem 2013). Bunching can also be used to overcome endogeneity in settings where the variable exhibiting bunching is the treatment, as recently shown by Caetano et al. (2020).

assumptions about heterogeneity typically used in the bunching design can also be relaxed. In the general formulation I consider, these take the form of assumptions about the marginal distributions of the two potential outcomes, which are observed in only a censored manner. These distributions must be extrapolated to estimate causal effects, as emphasized by Blomquist and Newey (2017) in the context of the isoelastic model. To perform this extrapolation I impose a weak nonparametric shape constraint—*bi-log-concavity*—on the distribution of each potential outcome. Bi-log-concavity nests many previously proposed assumptions for bunching analyses, and leads to a natural falsification test. The restriction affords partial identification of a conditional average treatment effect among individuals who in fact locate at the kink, a parameter I call the “buncher ATE”. In the overtime context, the buncher ATE represents an average reduced-form wage elasticity of hours demand, which I then use to assess the overall impact of the FLSA.

This result supplements other partial identification approaches recently proposed for the bunching design. Importantly, the bounds I derive for the buncher-ATE are substantially narrowed by making extrapolation assumptions separately for *each* of the two counterfactuals. By contrast, existing approaches operate by constraining the distribution of a single scalar heterogeneity parameter that appears in the typical isoelastic choice model. In the context of that model, Bertanha et al. (2020) and Blomquist et al. (2021) obtain bounds on the elasticity when the researcher is willing to put an explicit limit on how quickly the density of heterogeneous choices can change. My approach based on bi-log-concavity avoids the need to choose any such tuning parameters, in addition to being applicable in the general choice model.

The empirical setting of overtime pay involves two further challenges that are not typical to bunching-design analyses. Firstly, 40 hours is not an “arbitrary” point and bunching there could arise in part from factors other than it being the location of the kink. I use two strategies to estimate the amount of bunching that would exist at 40 absent the FLSA, and deliver clean estimates of the rule’s causal effect. My preferred strategy uses the fact that paid-time off hours do not count towards a week’s overtime threshold, shifting the location of the kink in a plausibly idiosyncratic way. A second consideration in applying the bunching design to overtime is that work hours may not always be set unilaterally by one party—in principle either the firm or the worker could have control over a given worker’s schedule, or the two might bargain over hours. I provide evidence that week-by-week variation in hours tends to be driven by the firm rather than the worker, and outline a conceptual framework to motivate this observation. This gets us into the conventional bunching design setup in which decisions are made by a single party, but I extend the results to a model in which bargaining weight between workers and firms is arbitrary and heterogeneous.

I find that the FLSA overtime rule does in fact reduce hours of work among hourly workers, despite there being reason to doubt that it would ultimately have any such effect (Trejo, 1991). My preferred estimate suggests that about one quarter of the bunching observed at 40 among hourly

workers is due to the FLSA, and those working at least 40 hours work, on average, about 30 minutes less than they would absent the time-and-a-half rule. Across specifications I estimate that the local wage elasticity of hours demand close to 40 falls in the range  $-0.04$  to  $-0.19$ . While the bunching design is only directly informative about the hours effects of the overtime rule—and not its employment effects—a back-of-the-envelope calculation using this estimate suggests that FLSA regulation creates about 700,000 jobs. I also use these estimates to evaluate hypothetical reforms to the FLSA, showing that even my generalized bunching-design model can be informative about counterfactual policies that change the location or “sharpness” of the kink.

The structure of the paper is as follows. Section 2 lays out a motivating conceptual framework for work hours that relates my approach to existing literature on overtime. Section 3 introduces the payroll data I use in the empirical analysis. In Section 4 I develop the generalized bunching-design approach, with Appendix A developing some of the supporting formal results. Section 5 applies these results to estimate effect of the FLSA overtime rule on hours worked, as well as the effects of hypothetical reforms to the FLSA. Section 6 discusses the empirical findings from the standpoint of policy objectives, and 7 concludes.

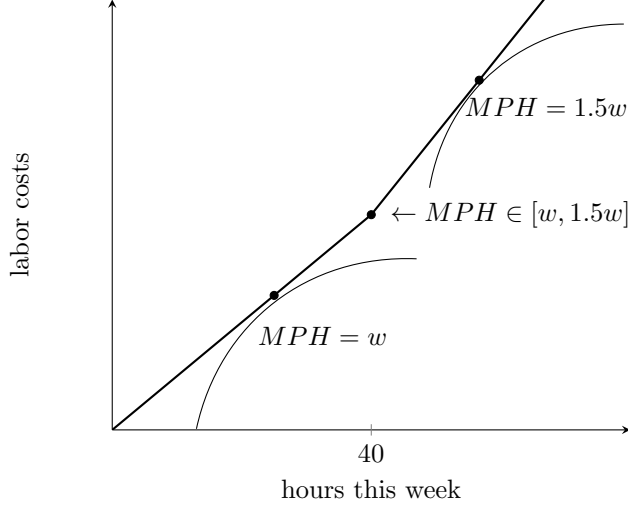
## 2 Conceptual framework

This section outlines a framework for thinking about the determination of weekly hours, which then motivates the identification strategy of Section 4. Readers primarily interested in the bunching design may wish to skip directly to that section.

The conceptual framework is centered around two observations from the data: weekly hours vary considerably between pay periods for an individual hourly worker, and a given worker’s hourly wage tends to change infrequently. I propose to view this as a two stage-process. First, workers are hired with an hourly wage set along with an “anticipated” number of hours they will work per week. Then, with that hourly wage fixed in the short-run, final scheduling of hours is controlled by the firm and varies by week given shocks to the firm’s demand for labor. Given the FLSA overtime rule and a worker’s fixed wage, their employer thus faces a kinked cost schedule when choosing hours in a given week, as pictured in Figure 1.

### Wages and anticipated hours set at hiring

We begin with the hiring stage, which pins down a worker’s wage. Following the literature I refer to the hourly rate of pay that applies to the first 40 of a worker’s hours as their *straight-time wage* or simply *straight wage*. The following discusses a benchmark model to endogenize these straight wages, which is spelled out formally in Appendix E. The bunching design strategy of Section 4 only



**FIGURE 1:** With a given worker's straight-time wage fixed at  $w$ , labor costs as a function of hours have a convex kink at 40 hours, given the overtime rule. Simple models of week-by-week hours choice (see Section 4.2) yield bunching when the marginal product of an hour at 40 is between  $w$  and  $1.5w$  for a mass of workers.

requires that *some* straight-time wage is agreed upon and fixed in the short-run for each worker, as is indeed observed in the data.

Suppose that firms hire by posting an earnings-hours pair  $(z, h)$ , where  $z$  is total weekly compensation offered to each worker, and  $h$  is the number of hours they are each expected to work per week. The firm faces a labor supply function  $N(z, h)$  determined by workers' preferences over the labor-leisure tradeoff,<sup>6</sup> and makes a choice of  $(z^*, h^*)$  given this labor supply function and their production technology. For simplicity, workers are here taken to be homogeneous in production, all paid hourly, and all covered by the FLSA.

While labor supply is viewed as a function over *total* compensation  $z$  and hours, there is always a unique straight wage associated with a particular  $(z, h)$  pair, such that  $h$  hours of work yields earnings of  $z$ , given the FLSA overtime rule:

$$w_s(z, h) = \frac{z}{h + \mathbb{1}(h > 40)0.5(h - 40)} \quad (1)$$

We can distinguish between the two main views on the likely effects of overtime policy by supposing that a workers' straight-time wage is set according to Eq. (1), given values  $z^*$  and  $h^*$  that the firm and worker agree upon at the time of hiring. Trejo (1991) calls these two views the *fixed-job* and the *fixed-wage* models of overtime.

The *fixed-job* view observes that for a generic smooth labor supply function  $N(z, h)$  (and rev-

<sup>6</sup>This labor supply function can be viewed as an equilibrium object that reflects both worker preferences and the competitive environment for labor. In the Supplemental Material, I endogenize this function in a simple extension of the imperfectly competitive Burdett and Mortensen (1998) search model.

enue production function with respect to hours), the optimal job package  $(z^*, h^*)$  for the firm to post will be the same as the optimal one absent the FLSA, as the hourly wage rate simply adjusts to fully neutralize the overtime premium.<sup>7</sup> Suppose for the moment that all workers in fact worked exactly  $h^*$  hours in all weeks (abstracting away from any reasons for the firm to deviate from  $h^*$  in any given week). Then the FLSA would have no effect on earnings, hours or employment, provided that  $w_s(z^*, h^*)$  is above any applicable minimum wage.

In the *fixed-wage* view, the firm faces an exogenous straight-time wage when determining hours. Versions of this idea are considered in Brechling (1965), Rosen (1968), Ehrenberg (1971), Hamermesh (1996), Hart (2004) and Cahuc and Zylberberg (2004). In a static model, this can be captured by a discontinuous labor supply function  $N(z, h)$ : one that exhibits perfect competition on the quantity  $w_s(z, h)$ . I show in Appendix E that in this case  $h^*$  and  $z^*$  are pinned down by the concavity of production with respect to hours and the scale of fixed costs (e.g. training) that do not depend on hours. The fixed-wage job makes the clear prediction that the FLSA will cause a reduction in hours, and bunching at 40.<sup>9</sup>

Trejo (1991) and Barkume (2010) investigate whether the fixed-job or fixed-wage model better accords with the observed joint distribution of hourly wages and hours. They find that wages do tend to be lower among jobs with overtime pay provisions and more overtime hours, but by a magnitude smaller than would be predicted by the fixed-jobs model. However, these estimates could be driven by selection, e.g. of lower-skilled workers into covered jobs with longer hours. In Appendix F, I construct an empirical test of Equation (1) that is instead based on assuming that the conditional distribution (across individual paychecks) of  $z$  is smooth across  $h = 40$ . I find that roughly one quarter of paychecks around 40 hours reflect the wage/hours relationship predicted by the fixed-job model.

This finding, along with the observation that hours change much more frequently than wages, is consistent with a model in which we imagine that after straight-wages are set according to Equation (1), they remain fairly static over time.<sup>10</sup> In common with the fixed wage model, this two-stage framework allows for the possibility that the overtime rule affects hours, and predicts bunching at 40. However, this is driven by short-run rigidity in straight-wages, rather than by perfect competition.

---

<sup>7</sup>In Appendix E I give a closed-form expression for  $(z^*, h^*)$  when both labor supply and production are iso-elastic: hours and earnings are each<sup>8</sup> increasing in the elasticity of labor supply with respect to earnings, and decreasing in the magnitude of the elasticity of labor supply with respect to pay.

<sup>9</sup>A fixed-wage model tends to predict an overall positive effect on employment given plausible assumptions on substitution between labor and capital (Cahuc and Zylberberg, 2004), though the total number of labor-hours will decrease (Hamermesh, 1996).

<sup>10</sup>This dovetails other recent evidence of uniformity and discretion in wage-setting, e.g. nominal wage rigidity (Grigsby et al. 2020), wage standardization (Hjort et al., 2020) and bunching at round numbers (Dube et al., 2020).

### Dynamic adjustment to hours by week

Confronted with the observation that workers' hours vary considerably week-to-week in my sample of hourly workers, I assume that this week-level variation reflects choices made by their employers. There are many reasons to expect variation in firms' demand for hours over time. Shocks to product demand or productivity change the number of weekly hours that would be optimal that week from the firm's perspective. If demand for the firm's products is seasonal or volatile, it may not be worthwhile to hire additional workers only to reduce employment later. Similarly, variation in productivity across workers may only become apparent to supervisors after their straight wages have been set, and it may be profitable to increase the hours of the most productive workers.

Throughout Section 4, I maintain a strong version of the assumption that a firm—rather than a worker—chooses the hours I observe on each paycheck. Workers' preferences *do* matter in the determination of each worker's straight wage, but I assume the firm has final scheduling rights week by week.<sup>11</sup> This eases notation and emphasizes the intuition behind my identification strategy. Appendix C presents a generalization in which some fraction of workers choose their hours, along with intermediate cases in which the firm and worker bargain over hours each week. If some but not all workers have full control of their weekly hours, then the bunching-design strategy will only be informative about effects of the FLSA among workers whose final hours are chosen by the firm.

Available survey evidence suggests that this group is the dominant one: a relatively small share of workers report that they choose their own schedules (despite the ongoing increase in flexible work arrangements). For example, the 2017-2018 Job Flexibilities and Work Schedules Supplement of the American Time Use Survey asks workers whether they have some input into their schedule, or whether their firm decides it. Only 17% report that they have some input. In a survey of firms, about 10% report that most of their employees have control over their shifts (Society for Human Resource Management, 2018).

## 3 Data and descriptive patterns

The main dataset I use comes from a large payroll processing company. They provided anonymized paychecks for the employees of 10,000 randomly sampled employers, for all pay periods in the years 2016 and 2017. At the paycheck level, I observe the check date, straight wage, and amount of pay and hours corresponding to itemized pay types, including normal (straight-time) pay, overtime pay, sick leave, holiday pay, and paid time off. The data also include state and industry for each employer and for employees: age, tenure, gender, state of residence, pay frequency and their salary

---

<sup>11</sup>This can be rationalized on the basis of workers generally having less bargaining power: if the worker and firm fail to agree on a worker's hours, the worker's outside option may be unemployment while the firm's outside option is having one less worker (Stole and Zwiebel, 1996).



if one is stored in the system.

### 3.1 Sample description

I construct a final sample for analysis based on two desiderata: a) the ability to observe hours within a single week; and b) a sample only of workers who are not exempt from the FLSA overtime rule. For the purposes of a), I keep paychecks from workers who are paid on a weekly basis (roughly half of the workers in the sample). To achieve b) I focus on hourly workers, since nearly all workers who are paid hourly are subject to FLSA regulation. The final sample includes 630,217 paychecks for 12,488 workers across 566 firms.<sup>12</sup>

Table 1 shows how the sample compares to survey data that is constructed to be representative of the U.S. labor force. Column (1) reports means from the final sample used in estimation, while (2) reports means before sampling restrictions. Column (3) reports means from the Current Population Survey (CPS) for the same years 2016–2017, among individuals reporting hourly employment. The “gets overtime” variable for the CPS sample indicates that the worker usually receives overtime, tips, or commissions. Column (4) reports means for 2016–2017 from the National Compensation Survey (NCS), a representative establishment-level dataset accessed on a restricted basis from the Bureau of Labor Statistics. The NCS reports typical overtime worked at the quarterly level for each job in an establishment, drawn from administrative data when possible.<sup>13</sup>

	(1)	(2)	(3)	(4)
	Estimation sample	Initial sample	CPS	NCS
Tenure (years)	3.21	2.81	6.34	.
Age (years)	37.15	35.89	39.58	.
Female	0.23	0.46	0.50	.
Weekly hours	38.92	27.28	36.31	35.70
Gets overtime	1.00	0.37	0.17	0.52
Straight-time wage	16.16	22.17	18.09	23.31
Weekly overtime hours	3.56	0.94	.	1.04
Number of workers in sample	12488	149459	63404	228773

**TABLE 1:** Comparison of the sample with representative surveys. Columns 1 and 2 average across periods within worker from the administrative payroll sample, and then present means across workers. Column 2 presents means of worker-level data from the Current Population Survey and Column 3 averages representative job-level data from the National Compensation Survey.

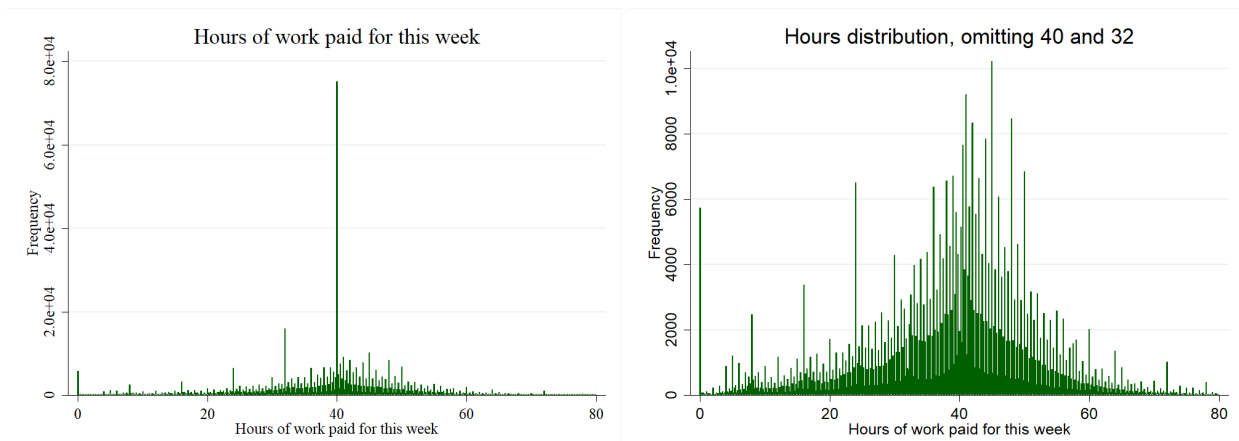
<sup>12</sup>Further details of the sample construction is given in Appendix F.1

<sup>13</sup>The hourly wage variable for the CPS may mix straight-time and overtime rates, and is only present in the outgoing rotation group sample. The tenure variable comes from the 2018 Job Tenure Supplement. The NCS does not distinguish between hourly and salaried workers, reporting an average hourly rate that includes salaried workers, who tend to be paid more. This likely explains the higher value than the CPS and payroll samples.

The sample I use is somewhat more male, earns lower straight-time wages, and works more overtime than a typical U.S. worker. Column (2) in Table 1 reveals that my sampling restrictions can explain why the estimation sample tilts male and has higher overtime hours than the workforce as a whole. In particular, conditioning on workers that are paid on a weekly basis oversamples industries that tend to have more men, and tend to pay somewhat lower wages. Appendix F compares the industry and regional distributions of the estimation sample to the CPS.

### 3.2 Hours and wages in the sample

I turn now to the main variables to be used in the analysis. Figure 2 reports the distribution of hours of work in the pooled sample of paychecks. The graphs indicate a large mass of individuals who were paid for exactly 40 hours that week, amounting to about 11.6% of the sample.<sup>14</sup> Appendix Figure F.13 shows that overtime pay is present in nearly all weekly paychecks that report more than 40 hours, in line with the presumption that workers in the final sample are not FLSA-exempt.



**FIGURE 2:** Empirical densities of hours worked pooling all paychecks in final estimation sample. Sample is restricted to hourly workers receiving overtime pay at some point (to ensure nearly all are non-exempt from FLSA, see text), and workers having hours variation. The right panel omits the points 40 and 32 to improve visibility elsewhere. Bins have a width of 1/8 of an hour, based on the observed granularity of hours (see Appendix Figure 4 for details).

Table 2 documents that while the hours paid in 70% of all pay checks in the final estimation sample differ from those of the last paycheck by at least one hour, just 4% of all paychecks record a different straight-time wage than the previous paycheck for the same worker. Among the roughly 22,500 wage change events, the average change is about a 45 cent raise per hour. When hours change the magnitude is about 7 hours on average (see Supplemental Figure 5 for the distribution of hours changes), with no average secular increase in hours over time.

<sup>14</sup>The second largest mass occurs at 32 hours, and is explained by paid-time-off, holiday, and sick pay hours as discussed in Section 5.

	Mean	Std. dev.	N
Indicator for hours changed from last period	0.84	0.37	630,217.00
Indicator for hours changed by at least 1 hour	0.70	0.46	630,217.00
Indicator for wage changed from last period	0.04	0.19	630,217.00
Indicator for wage changed, if hours changed	0.04	0.19	529,791.00
Difference in hours, if hours changed	-0.02	10.69	529,791.00
Absolute value of hours difference, if hours changed	6.83	8.23	529,791.00
Difference in wage, if wage changed	0.45	26.46	22,501.00

**TABLE 2:** Changes in hours or straight wages between a worker’s consecutive paychecks.

Appendix F reports some further details on the variation in hours and wages present in the data. Appendix Table F.3 regresses hours, overtime hours, and an indicator for bunching on worker observables, and shows that after controlling for worker and date fixed effects bunching and overtime hours are both predicted by recent hiring at the firm, lending further evidence for the assumption that shocks to labor demand drive variation in hours. Appendix Table F.4 shows that overall, about 63% of variation in total hours can be explained by worker and employer-by-date fixed effects. Appendix Figure F.2 documents heterogeneity in the prevalence of overtime pay across industry classifications. Appendix Table F.6 reports a direct estimate of the proportion of wage-hours pairs that are related according to Equation (1), close to 40.

## 4 Empirical strategy: a generalized kink bunching design

Let us now turn to the firm choosing the hours of a given worker in a particular week, with that worker’s wage fixed and costs a kinked function of hours as depicted in Figure 1. We will see that under weak assumptions, firms facing such kinks will make choices that can be completely characterized by choices they *would* make under two counterfactual linear cost schedules that differ with respect to wage. I relate the observable bunching at 40 hours to a treatment effect defined from these two counterfactuals, which I then use to estimate the impact of the FLSA on hours.

The identification results in this section hold in a much more general setting in which a generic decision-maker faces a generic “kinked” choice set and has convex preferences. I present this general model in Appendix A. Throughout this section I refer to a worker  $i$  in week  $t$  as a *unit*: an observation of  $h_{it}$  for unit  $it$  is thus the hours recorded on a single paycheck. Probability statements are to be understood with respect to such paycheck-level units.

## 4.1 A general choice model

Let us start from the conceptual framework introduced in Section 2. In choosing the hours  $h_{it}$  of worker  $i$  in week  $t$ , worker  $i$ 's employer faces a kinked cost schedule, given the worker's straight-time wage this week  $w_{it}$ .<sup>15</sup> If the firm chooses less than 40 hours, it will pay  $w = w_{it}$  for each hour, and if the firm chooses  $h > 40$  it will pay  $40w$  for the first 40 hours and  $1.5w(h - 40)$  for the remaining hours, giving the convex shape to Figure 1. We can write the kinked pay schedule for unit  $it$ , as a function of hours this week  $h$ , as

$$B_{kit}(h) = w_{it}h + .5w_{it}\mathbb{1}(h > 40)(h - 40) = \max\{B_{0it}(h), B_{1it}(h)\}$$

where  $B_{0it}(h) = w_{it}h$  and  $B_{1it}(h) = 1.5w_{it}h - 20w_{it}$ . The kinked pay schedule  $B_{kit}(h)$  is equal to  $B_{0it}(h)$  for values  $h \leq 40$  and  $B_{kit}(h)$  is equal to  $B_{1it}(h)$  for values  $h \geq 40$ . The functions  $B_0$  and  $B_1$  recover the two segments in Figure 1, and are depicted separately in Appendix Figure A.1.

**Definition (potential outcomes).** Let  $h_{0it}$  denote the hours of work that of unit  $it$  would be paid for if instead of  $B_{kit}(h)$ , the pay schedule for this week's hours were  $B_{0it}(h)$ . Similarly, let  $h_{1it}$  denote the hours of pay that would occur for unit  $it$  if the pay schedule were  $B_{1it}(h)$ .

By contrast, let  $h_{it}$  denote the actual hours of work for which unit  $it$  is paid. Our first assumption is that actual hours and potential outcomes reflect choices made by the firm:

**Assumption CHOICE.** Each of  $h_{0it}$ ,  $h_{1it}$  and  $h_{it}$  reflect choices the firm would make under pay schedules of the form  $z = B(h)$ , with  $B(h)$  given by  $B_{0it}(h)$ ,  $B_{1it}(h)$ , and  $B_{kit}(h)$  respectively.

CHOICE simply reflects the assumption that hours are perfectly manipulable by firms. Note that if firm preferences over a unit's hours are quasi-linear with respect to costs (e.g. if they maximize weekly profits), the term  $-20w_{it}$  appearing in  $B_{1it}$  plays no role in firm choices. As such, I will often refer to  $h_{1it}$  as choice made under linear pay at the overtime rate  $1.5w_{it}$ , keeping in mind that the definition given above is necessary for the analysis if preferences are not quasi-linear.

Our second assumption is that each unit's firm optimizes some vector  $\mathbf{x}$  of choice variables that pin down that unit's hours. As a leading case, we may think of hours of work as a single component of firms' choice vector  $\mathbf{x}$  (Section A.3 gives some examples). Firm preferences are taken to be convex in  $\mathbf{x}$  and the unit's wage costs  $z$ :

**Assumption CONVEX.** Firm choices for unit  $it$  maximize some  $\pi_{it}(z, \mathbf{x})$ , where  $\pi_{it}$  is strictly quasiconcave in  $(z, \mathbf{x})$  and decreasing in  $z$ . Hours are a continuous deterministic function of  $\mathbf{x}$ .

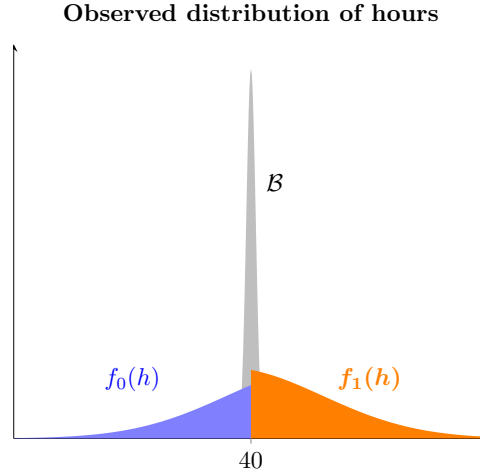
<sup>15</sup>A unit's straight-time wage  $w_{it}$  is fixed with respect to the choice of hours this week, but may depend on  $t$  due to e.g. occasional or automatic periodic raises.

That firms rather than workers choose hours enters in the assumption that  $\pi$  is decreasing (rather than increasing) in  $z$ , but Appendix C relaxes this to allow some workers to set their hours. Appendix A shows that bunching still has some identifying power without convexity of preferences.

The starting point for our analysis of identification in the bunching design is the following mapping between actual hours  $h_{it}$  and the counterfactual hours choices  $h_{0it}$  and  $h_{1it}$ . Appendix Lemma 1 shows that Assumptions CHOICE AND CONVEX imply that:

$$h_{it} = \begin{cases} h_{0it} & \text{if } h_{0it} < 40 \\ 40 & \text{if } h_{1it} \leq 40 \leq h_{0it} \\ h_{1it} & \text{if } h_{1it} > 40 \end{cases} \quad (2)$$

That is, a worker will work  $h_{0it}$  hours when the counterfactual choice  $h_{0it}$  is less than 40, and  $h_{1it}$  hours when  $h_{1it}$  is greater than 40. They will be located at the corner solution of 40 if and only if the two counterfactual outcomes fall on either side, “straddling” the kink.<sup>16</sup> Figure 3 depicts the implications of Eq. (2) for what is observable by the researcher in the bunching design: censored distributions of both  $h_0$  and of  $h_1$ , and a point-mass of size  $\mathcal{B} = P(h_{1it} \leq 40 \leq h_{0it})$  at the kink.



**FIGURE 3:** Observables in the bunching design, given Equation (2). To the left of the kink at 40, the researcher observes the density  $f_0(h)$  of the counterfactual  $h_{0it}$ , up to values  $h = 40$ . To the right of the kink, the researcher observes the density  $f_1(h)$  of  $h_{1it}$  for values  $h > 40$ . At the kink, one observes a point-mass of size  $\mathcal{B} = P(h_{1it} \leq 40 \leq h_{0it})$ .

Equation (2) represents a departure from previous approaches to the bunching design, which have characterized bunching in terms of the counterfactual  $h_0$  only. We will see that this is a simplifica-

<sup>16</sup>“Straddling” can only occur in one direction, with  $h_{1it} \leq k \leq h_{0it}$ . The other direction:  $h_{0it} \leq k \leq h_{1it}$  with at least one inequality strict, is ruled out by the weak axiom of revealed preference (see Appendix A).

tion afforded by the benchmark isoelastic utility model, but in a generic choice model, both  $h_0$  and  $h_1$  are necessary to pin down actual choices  $h_{it}$ . Appendix A shows that Eq. (2) also holds in settings with possibly non piecewise-linear kinked choice sets of the form:  $z \geq \max\{B_0(\mathbf{x}), B_1(\mathbf{x})\}$  where  $B_0$  and  $B_1$  are weakly convex in the full vector  $\mathbf{x}$ , and  $z$  is any “cost” disliked by decision-makers.

### Intuition for Equation (2)

As an intuitive illustration of Equation (2), suppose that firms balance the cost  $B_{kit}(h)$  against the value of  $h$  hours of the worker’s labor, in order to maximize that week’s profits. Then Eq. (2) can be written:

$$h_{it} = \begin{cases} MPH_{it}^{-1}(w_{it}) & \text{if } MPH_{it}(40) < w_{it} \\ 40 & \text{if } MPH_{it}(40) \in [w_{it}, 1.5w_{it}] \\ MPH_{it}^{-1}(1.5w_{it}) & \text{if } MPH_{it}(40) > 1.5w_{it} \end{cases} \quad (3)$$

where denotes  $MPH_{it}(h)$  is the marginal product of an hour of labor for unit  $it$ , as a function of hours  $h$ . Assuming that production is strictly concave, the function  $MPH_{it}(h)$  will be strictly decreasing in  $h$ , and we have that  $h_{0it} = MPH_{it}^{-1}(w_{it})$  and  $h_{1it} = MPH_{it}^{-1}(1.5w_{it})$ .

Figure 1 depicts Eq. (3) visually. Consider for example a worker with a straight-wage of \$10 an hour. If there exists a value  $h < 40$  such that the worker’s  $MPH$  is equal to \$10, then the firm will choose this point of tangency. This happens if and only if the marginal product of an hour at 40 hours is less than \$10. If instead, the marginal product of an hour is still greater than \$15 at  $h = 40$ , the firm will choose the value  $h > 40$  such that  $MPH$  equals \$15. The third possibility is that the  $MPH$  at  $h = 40$  is *between* the straight and overtime rates \$10 and \$15. In this case, the firm will choose the corner solution  $h = 40$ , contributing to bunching at the kink.

While Eq. (3) provides a natural nonparametric characterization of when the firm will ask a worker to work overtime (when the ratio of productivity to wages is high), it is still more restrictive than necessary for the purposes of the bunching design. Appendix A.3 provides some examples that use the full generality of CONVEX, in which firms simultaneously consider *multiple* margins of choice aside from a given unit’s hours. For example, the firm may attempt to mitigate the added cost of overtime by reducing bonuses when a worker works many overtime hours. Eq. (2) remains valid even when such additional margins of choice are unmodeled and unobserved by the econometrician, varying possibly by unit.

Note that if production depends jointly on the hours of all workers within a firm, we may expect the function  $MPH_{it}(h)$  in Eq. (3) to depend on the hours of worker  $i$ ’s colleagues in week  $t$ . In this case the quantities  $h_{0it}$  and  $h_{1it}$  hold the hours of  $i$ ’s colleagues fixed at their *realized* values: they contemplate counterfactuals in which the pay schedule for a single unit  $it$  is varied,

and nothing else. This affects the interpretation of our treatment effects, as discussed in Section 4.4. In the baseline isoelastic model that we consider in the next section, such interdependencies between workers' hours are ruled out by virtue of assuming that production is linearly separable across units. See Appendix D for discussion of this setting with a general non-separable production function.

## 4.2 The benchmark iso-elastic model

The canonical approach from the bunching-design literature (Saez, 2010; Chetty et al., 2011; Kleven, 2016), strengthens Assumption CONVEX to suppose that  $\mathbf{x} = h$  and decision-makers' utility features an iso-elastic functional form, with preferences common between units up to a scalar heterogeneity parameter. This corresponds to a situation in which firm profits from unit  $it$  take the form:

$$\pi_{it}(z, h) = a_{it} \cdot \frac{h^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}} - z \quad (4)$$

where  $\epsilon < 0$  is common across units, and  $z$  represents wage costs for worker  $i$  in week  $t$ . This corresponds to the situation described in Eq. (3), in which  $MPH_{it}(h)$  depends on  $it$  only through the heterogeneity parameter  $a_{it}$ , which reflects a productivity shock for that unit. Eq. (4) is analogous to the isoelastic, quasilinear labor *supply* model used in the context of tax kinks.

Under a linear pay schedule  $z = wh$ , the profit maximizing number of hours is  $\left(\frac{w}{a_{it}}\right)^\epsilon$ , so  $\epsilon$  yields the elasticity of hours demand with respect to a linear wage. Let  $\eta_{it} = a_{it}/w_{it}$  denote the ratio of a unit's current productivity factor to their straight wage. In the iso-elastic model  $h_{0it} = \eta_{it}^{-\epsilon}$  and  $h_{1it} = 1.5^\epsilon \cdot \eta_{it}^{-\epsilon}$ , and by Eq. (3) actual hours  $h_{it}$  are ranked across units in order of  $\eta_{it}$ . In particular  $h_{0it} < 40$  whenever  $\eta_{it} < 40^{-1/\epsilon}$ ,  $h_{1it} > 40$  whenever  $\eta_{it} > 1.5 \cdot 40^{-1/\epsilon}$  and  $h_{1it} \leq 40 \leq h_{0it}$  when  $\eta_{it}$  falls in the intermediate range  $[40^{-1/\epsilon}, 1.5 \cdot 40^{-1/\epsilon}]$ . Whether a worker has overtime hours in a given week is determined by the ratio  $\eta_{it}$ : a worker with a wage  $w_{it}$  fixed throughout the year may for example work overtime in periods when  $a_{it}$  is relatively high due to seasonally elevated productivity. If  $\eta_{it}$  is continuously distributed over a region containing the bunching interval  $[40^{-1/\epsilon}, 1.5 \cdot 40^{-1/\epsilon}]$ , then the observed distribution of  $h_{it}$  will feature a point mass at 40—"bunching"—and a density elsewhere.

### Identification in the iso-elastic model

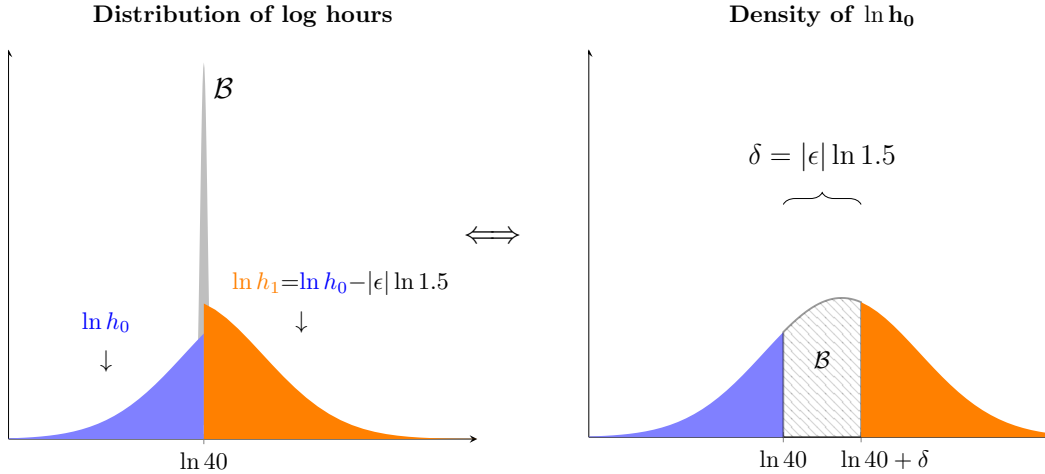
In the context of the iso-elastic model, a natural starting place for evaluating the FLSA is to estimate the parameter  $\epsilon$ . If we ignore any effects of the policy on straight-wages, the effect of the time-and-a-half rule on unit  $it$ 's hours will simply be the difference  $h_{it} - h_{0it}$ , what we might call the *effect of the kink*. It follows from the above that whenever  $h_{it} > 40$ , the effect of the kink is  $h_{it} \cdot (1 - 1.5^{-\epsilon})$ .

Provided the value of  $\epsilon$ , we could thus evaluate the effect of the kink for any paycheck recording overtime using that unit's observed hours.

The classic bunching-design method pioneered by Saez (2010) identifies  $\epsilon$  by relating it to the observable bunching probability:

$$\mathcal{B} = P(h_{it} = 40) = \int_{40}^{1.5^{|\epsilon|} \cdot 40} f_0(h) \cdot dh \quad (5)$$

where  $f_0$  is the density of  $h_0$ . If the function  $f_0$  were known, the value of  $\epsilon$  could be pinned down in a straightforward way from Eq. (5). Figure 4 depicts the intuition. Since  $h_{1it} = 1.5^\epsilon \cdot h_{0it}$  in the iso-elastic model, it is convenient to visualize this after applying a log transformation to hours. Treatment effects in logs,  $\delta = \ln h_{0it} - \ln h_{1it} = |\epsilon| \cdot \ln 1.5$ , are homogeneous across all units  $it$ , and the density of  $\ln h_{1it}$  is thus a simple leftward shift of the density of  $\ln h_{0it}$ , by  $\delta$ .



**FIGURE 4:** The left panel depicts the distribution of observed log hours  $\ln h_{it}$  in the iso-elastic model, while the right panel depicts the underlying full density of  $\ln h_{0it}$ . Specializing from the general setting of Figure 3, we have in the iso-elastic model that  $f_1(h) = f_0(h + |\epsilon| \cdot \ln 1.5)$ . Thus, the full density of  $f_0$  is related to the observed distribution by “sliding” the observed distribution for  $h > 40$  out by the unknown distance  $\delta = |\epsilon| \ln 1.5$ , leaving a missing region in which  $f_0$  is unobserved. The total area in the missing region from  $\ln 40$  to  $\ln 40 + \delta$  must equal the observed bunching mass  $\mathcal{B}$ .

Standard approaches in the bunching design make parametric assumptions about  $f_0$  to point-identify  $\epsilon$ .<sup>17</sup> The approach of Saez (2010), for example, assumes for example that the density of  $h_0$  is linear through the missing region of Figure 4. The popular method of Chetty et al. (2011) instead fits a global polynomial, using the distribution of hours outside the missing region to impute the density of  $h_0$  within it. Neither approach is particularly suitable in the overtime context. The linear method of Saez (2010) implies monotonicity of the density in the missing region, which is unlikely

<sup>17</sup>(Bertanha et al., 2020) note that given a full parametric model for  $f_0$ , the entire model could be estimated by maximum likelihood. This approach would enforce (5) automatically while enjoying the efficiency properties of MLE.



to hold given that 40 appears to be near the mode of the  $h_0$  latent hours distribution. The method of Chetty et al. (2011) ignores the “shift” by  $\delta$  in the right panel of Figure 4.<sup>18</sup>

If on the other hand, the researcher is unwilling to assume anything about the density of  $h_0$  in the missing region of Figure 4, then the data are compatible with any finite  $\epsilon < 0$ , a point emphasized by Blomquist et al. (2021) and Bertanha et al. (2020). In particular, given the integration constraint (5), an arbitrarily small  $|\epsilon|$  could be rationalized by a density that spikes sufficiently high just to the right of 40, while an arbitrarily large  $|\epsilon|$  can be reconciled with the data by supposing that the density of  $h_0$  drops quickly to some very small level throughout the missing region. I find a middle ground by imposing a nonparametric shape constraint on  $h_0$ : *bi-log-concavity* (BLC), leading to a partial identification result. A detailed discussion of BLC is given in Section 4.3.

### Limitations of the iso-elastic model

Compared with the iso-elastic model, the general choice model from Section 4.1 allows a wide range of underlying structural choice models that might drive a firm’s hours response to the FLSA. This robustness over structural models is important in the overtime context. As reported in Appendix F.5, assuming either that  $h_0$  is BLC or using the linear density method of Saez (2010) suggests that  $\epsilon \approx -0.2$ .<sup>19</sup> This is implausible when interpreted through the lens of Equation (4): it implies an hours production function of  $f(h) = -\frac{1}{4}h^{-4}$  (up to an affine transformation), which features an unreasonable degree of concavity. Allowing a more general non-exponential production function  $f(h)$  (separable between units) is also not much help, as the standard bunching design approach then estimates an averaged local inverse elasticity of  $f(h)$  (see Appendix F.5). Put simply, the observed bunching is too small to be reconciled with a model in which  $\epsilon$  parameterizes the concavity of weekly production with respect to hours. The estimand of the bunching design should instead be interpreted as a *reduced-form* elasticity of the demand for hours, defined in terms of the counterfactuals  $h_0$  and  $h_1$ . In the general choice model of CONVEX, additional margins of response (such as those considered in Appendix A.3) can attenuate the labor demand response to overtime pay, beyond what would be predicted on the basis of productivity considerations alone.

## 4.3 Identifying treatment effects in the general choice model

In this section I turn to identification in the general choice model of Section 4.1. Without a single preference parameter like  $\epsilon$  that characterizes responsiveness to incentives for all units, we face

<sup>18</sup>This is perhaps less problematic in typical settings where the bunching is somewhat diffuse around the kink. However in my setting bunching is exact, and the slope of the density is far from zero near 40.

<sup>19</sup>This estimate is from the pooled sample across all industries, and attributes all of the bunching observed at 40 to the FLSA (as I do in Section 5.1) would only further reduce the magnitude of  $\epsilon$ . This is also not driven by aggregating industries: estimation by industry yields bounds on  $\epsilon$  ranging from  $-0.26$  to  $-0.06$ , which are similarly implausible as estimates of concavity of production.

the necessary question: what parameter of interest might be identifiable from the data without the restrictive iso-elastic model, but still help us to evaluate the effect of the FLSA on hours?

I refer to the difference  $\Delta_{it} := h_{0it} - h_{1it}$  between  $h_0$  and  $h_1$  as unit  $it$ 's *treatment effect*. Recall that  $h_0$  and  $h_1$  are interpreted as potential outcomes, indicating what *would* have happened had the firm faced either of two counterfactual pay schedules instead of the kink.  $\Delta_{it}$  thus represents the causal effect of a one-period 50% increase in worker  $i$ 's wage on their hours in week  $t$ : the difference between the hours that unit's firm would choose if the worker were paid at their straight-time rate versus at their overtime rate for all hours in that week. A unit's treatment effect can be contrasted with the "effect of the kink" quantity  $h_{it} - h_{0it}$  introduced before, but importantly the two are related: by Eq. (2) the effect of the kink is  $-\Delta_{it}$  for all units working overtime.

In the iso-elastic model  $\Delta_{it} = h_{0it} \cdot (1 - 1.5^\epsilon)$ , representing a special case in which treatment effects are constant across all units after a log transformation of the outcome:  $\ln h_{0it} - \ln h_{1it} = |\epsilon| \cdot \ln 1.5$ . In general we can expect  $\Delta_{it}$  to vary much more flexibly across units, and a reasonable parameter of interest becomes a summary statistic of  $\Delta_{it}$  of some kind. In particular, Eq. (2) suggests that bunching is informative about the distribution of  $\Delta_{it}$  among units "near" the kink. To see this, let  $k = 40$  denote the location of the kink, and write the bunching probability as:

$$\mathcal{B} = P(h_{1it} \leq k \leq h_{0it}) = P(h_{0it} \in [k, k + \Delta_{it}]) = P(h_{1it} \in [k - \Delta_{it}, k]), \quad (6)$$

i.e. units bunch when their  $h_0$  potential outcome lies to the right of the kink, but within that unit's individual treatment effect of it. Note that by Eq. (2) we can also write bunching in terms of the marginal distributions of  $h_0$  and  $h_1$ :  $\mathcal{B} = F_1(k) - F_0(k)$ , provided that they are continuously distributed and with  $F_0$  and  $F_1$  their cumulative distribution functions.

### Parameter of interest: the buncher ATE

I focus my identification analysis on the average treatment effect among units who locate at exactly 40 hours, a parameter I call the "buncher ATE". In the overtime setting some care is needed in defining this parameter, as recall that it natural to expect a mass of such buncher' at 40, even absent the FLSA. Let us indicate such "counterfactual bunchers" by an (latent) binary variable  $K_{it}^* = 1$ , and define the buncher ATE to be:

$$\Delta_k^* = \mathbb{E}[\Delta_{it} | h_{it} = k, K_{it}^* = 0],$$

That is,  $\Delta_k^*$  is the average value of  $\Delta_{it}$  among bunchers who bunch in response to the FLSA kink, and would not locate at 40 hours otherwise. In evaluating the FLSA, I will suppose that all counterfactual bunchers have a zero treatment effect, such that  $h_{0it} = h_{1it} = k$ . Since  $\Delta_{it} = 0$  for these

units by assumption, we can move between  $\Delta_k^*$  and  $\mathbb{E}[\Delta_{it}|h_{it} = k]$ , provided the counterfactual bunching mass  $p := P(K_{it}^* = 1)$  is known. In this section, I treat  $p$  as given, and estimate it empirically in Section 5.1.

To simplify the discussion, suppose for now that there are no counterfactual bunchers, so that  $\Delta_k^* = \mathbb{E}[\Delta_{it}|h_{it} = k]$ . Our goal is to invert (6) in some way to learn about the buncher ATE from the observable bunching probability  $\mathcal{B}$ . In Figure 4, we've seen the intuition for this exercise in the context iso-elastic model, in which there is only one dimension of heterogeneity and  $h_{1it} = h_{0it} \cdot 1.5^\epsilon$ . The key implication of the iso-elastic model that aids in identification is *rank invariance* between  $h_0$  and  $h_1$ . Rank invariance (Chernozhukov and Hansen 2005) says that  $F_0(h_{0it}) = F_1(h_{1it})$  for all units, i.e. increasing each unit's wage by 50% does not change their rank in the hours distribution. For example, a worker at the median of the  $h_0$  distribution also has a median value of  $h_1$ . Rank invariance is satisfied by models in which there is perfect positive co-dependence between the potential outcomes (see left panel of Figure 5).

Rank invariance is useful because it allows us to translate statements about  $\Delta_{it}$  into statements about the marginal distributions of  $h_{0it}$  and  $h_{1it}$ . In particular, under rank invariance the buncher ATE is equal to the quantile treatment effect  $Q_0(u) - Q_1(u)$  averaged across all  $u$  between  $F_0(k)$  and  $F_1(k) = F_0(k) + \mathcal{B}$ , with  $Q_d$  the quantile function of  $h_{dit}$ , i.e.:

$$\Delta_k^* = \frac{1}{\mathcal{B}} \int_{F_0(k)}^{F_1(k)} [Q_0(u) - Q_1(u)] du, \quad (7)$$

so long as  $F_0(y)$  and  $F_1(y)$  are continuous and strictly increasing. I focus on partial identification of the buncher ATE, for which it is sufficient to place point-wise bounds on the quantile functions  $Q_0(u)$  and  $Q_1(u)$  throughout the range  $u \in [F_0(k), F_1(k)]$  as depicted in Figure 6.

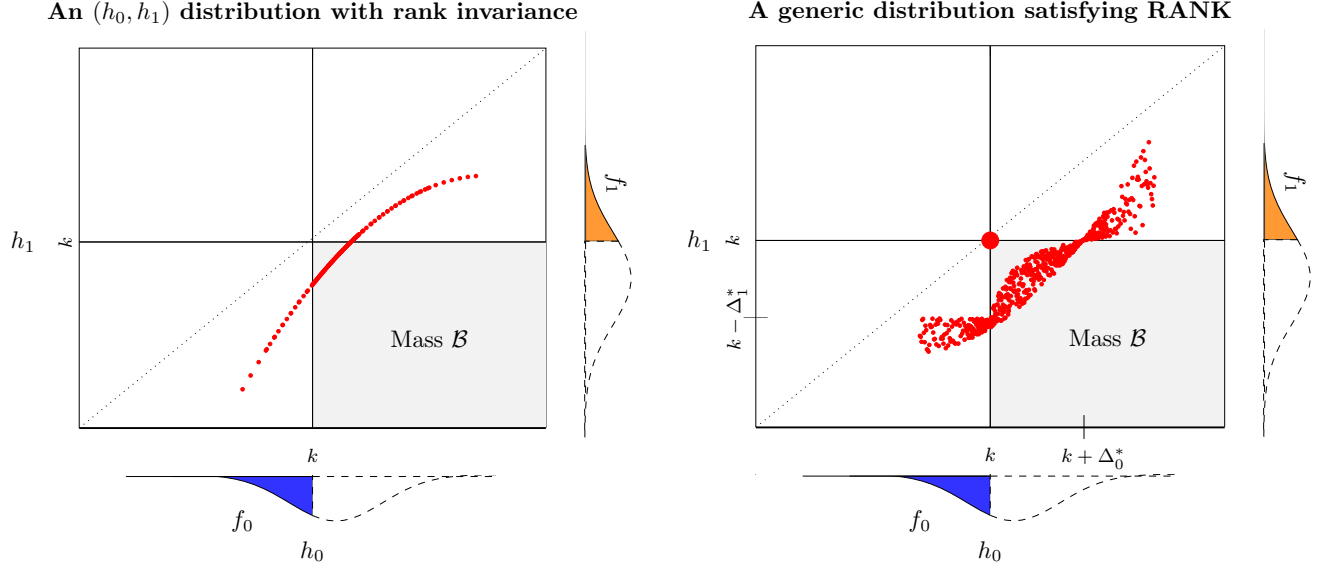
While rank invariance already relaxes the iso-elastic model used thus far in the literature, a still weaker assumption proves sufficient for Eq. (7) to hold:

**Assumption RANK.** *There exist fixed values  $\Delta_0^*$  and  $\Delta_1^*$  such that  $h_{0it} \in [k, k + \Delta_{it}]$  iff  $h_{0it} \in [k, k + \Delta_0^*]$ , and  $h_{1it} \in [k - \Delta_{it}, k]$  iff  $h_{1it} \in [k - \Delta_1^*, k]$ .*

Unlike (strict) rank invariance, Assumption RANK allows ranks to be reshuffled by treatment among bunchers and among the group of units that locates on each side of the kink.<sup>20</sup> For example, suppose that a 50% increase in the wage of worker  $i$  would result in their hours being reduced from  $h_{0it} = 50$  to  $h_{1it} = 45$ . If another worker  $j$ 's hours are instead reduced from  $h_{0jt} = 48$  to

<sup>20</sup>When  $p = 0$  Assumption RANK is equivalent to an instance of the *rank-similarity* assumption of Chernozhukov and Hansen (2005), in which the conditioning variable is which of the three cases of Equation (2) hold for the unit. Specifically, for both  $d = 0$  and  $d = 1$ :  $U_d|(h < k) \sim Unif[0, F_0(k)]$ ,  $U_d|(h = k) \sim Unif[F_0(k), F_1(k)]$ , and  $U_d|(h > k) \sim Unif[F_1(k), 1]$ .

$h_{1jt} = 46$  under a 50% wage increase, workers  $i$  and  $j$  will switch ranks, without violating RANK. Note also that RANK is compatible with the existence of counterfactual bunchers  $p > 0$ .



**FIGURE 5:** The joint distribution of  $(h_{0it}, h_{1it})$  (in red), comparing an example satisfying rank invariance (left) to a case satisfying Assumption RANK (right). RANK allows the support of the joint distribution to “fan-out” from perfect co-dependence of  $h_0$  and  $h_1$ , except when either outcome is equal to  $k$ . The large dot in the right panel indicates a possible mass  $p$  of counterfactual bunchers. The observable data identifies the shaded portions of outcome’s marginal distribution (depicted along the bottom and right edges), as well as the total mass  $\mathcal{B}$  in the (shaded) south-east quadrant.

The right panel of Figure 5 shows an example of a distribution satisfying RANK, which requires the support of  $(h_0, h_1)$  to narrow to a point as it crosses  $h_0 = k$  or  $h_1 = k$ . When this is not perfectly satisfied, Appendix Figure A.2 demonstrates how the RHS of Equation (7) will then yield a lower bound on the true buncher ATE (and can still be interpreted as an averaged quantile treatment effect). Appendix Figure C.4 generalizes RANK to case in which some workers choose their hours, resulting in mass in the north-west quadrant of Figure 5.

#### 4.3.1 Bounds on the buncher ATE via bi-log-concavity

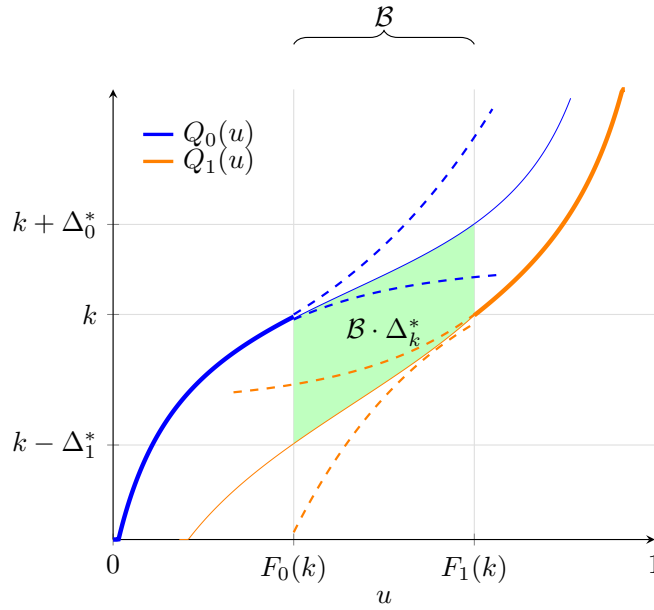
Given Eq. (7), I obtain bounds on the buncher ATE by assuming that both  $h_0$  and  $h_1$  have *bi-log-concave* distributions. Bi-log-concavity is a nonparametric shape constraint that generalizes log-concavity, a property of many common parametric distributions:

**Definition (BLC).** A distribution function  $G$  is *bi-log-concave (BLC)* if both  $\ln G$  and  $\ln(1 - G)$  are concave functions.

If  $G$  is BLC then it admits a strictly positive density  $g$  that is itself differentiable with the locally bounded derivative:  $\frac{-g(h)^2}{1-G(h)} \leq g'(h) \leq \frac{g(h)^2}{G(h)}$  (Dümbgen et al., 2017). Intuitively, this will rule out

cases in which the density of  $h_0$  or  $h_1$  ever spikes or falls *too* quickly on the interior of its support, leading to non-identification of the type discussed in Section 4.2.<sup>21</sup>

The family of BLC distributions includes parametric distributions assumed by previous bunching design studies, such as those with uniform or linear densities Saez (2010), or those with polynomial densities as in Chetty et al. 2011 (provided they have real roots). All globally log-concave distributions are BLC.<sup>22</sup> Importantly, the BLC property is partially testable in the bunching design, since  $F_0(y)$  is identified for all  $h < k$  and  $F_1(h)$  is identified for all  $h > k$ . Appendix Figure F.12 shows that the observable portions of  $F_0$  and  $F_1$  are indeed consistent with BLC, passing this falsification test. Identification requires us to believe that the unobserved portions of  $F_0$  and  $F_1$  also satisfy this weak shape constraint.



**FIGURE 6:** Extrapolating the quantile functions for  $h_0$  and  $h_1$  (blue and orange, respectively) to place bounds on the buncher ATE (case depicted has no counterfactual bunchers). The observed portions of each quantile function are depicted by thick curves, while the unobserved portions are indicated by thinner curves. The dashed curves represent upper and lower bounds for this unobserved portion coming from an assumption like bi-log-concavity (see text below). The buncher ATE is equal to the area shaded in green, divided by the bunching probability  $\mathcal{B}$ . The quantities  $\Delta_0^*$  and  $\Delta_1^*$  are defined in Assumption RANK below.

We are now ready to state the main identification result. Its logic is as follows: given the general choice model, RANK converts identification of the buncher ATE into a pair of extrapolation problems, each of which are approached by assuming bi-log-concavity of the corresponding marginal

<sup>21</sup>Bertanha et al. (2020) propose partial identification in an iso-elastic model by specifying a Lipschitz constant on the density of  $\ln \eta_{it}$ . This yields global rather than local bounds on  $g'$ .

<sup>22</sup>However unlike log-concave densities, BLC distributions need not be unimodal (Dümbgen et al., 2017).

potential outcome distribution.<sup>23</sup> Let  $F(h) := P(h_{it} \leq h)$  be the CDF of observed hours.

**Theorem 1 (bi-log-concavity bounds on the buncher ATE).** *Assume CHOICE, CONVEX, RANK and that  $h_{0it}$  and  $h_{1it}$  have bi-log-concave distributions conditional on  $K_{it}^* = 0$ . Then:*

1. *Each of  $F(h)$ ,  $F_0(h)$  and  $F_1(h)$  are continuously differentiable for  $h \neq k$ .  $F_0(k) = \lim_{h \uparrow k} F(h) + p$ ,  $F_1(k) = F(k)$ ,  $f_0(k) = \lim_{h \uparrow k} f(h)$  and  $f_1(k) = \lim_{h \downarrow k} f(h)$ , where we define the density  $f_d(y)$  of  $h_{dit}$  at  $y = k$  to be  $f_d(k) = \lim_{h \rightarrow k} f_d(h)$ , for each  $d \in \{0, 1\}$ , when  $p > 0$ .*

2. *The buncher ATE  $\Delta_k^*$  lies in the interval  $[\Delta_k^L, \Delta_k^U]$ , where:*

$$\Delta_k^L := g(F_0(k) - p, f_0(k), \mathcal{B} - p) + g(1 - F_1(k), f_1(k), \mathcal{B} - p)$$

and

$$\Delta_k^U := -g(1 - F_0(k), f_0(k), p - \mathcal{B}) - g(F_1(k) - p, f_1(k), p - \mathcal{B})$$

with  $g(a, b, x) = \frac{a}{bx} (a + x) \ln(1 + \frac{x}{a}) - \frac{a}{b}$ , and the bounds are sharp.

*Proof.* See Appendix B. □

Combining Items 1 and 2 of Theorem 1, it follows that the sharp bounds  $\Delta_k^L$  and  $\Delta_k^U$  on the buncher ATE are identified, given the CDF of the data  $F(h)$  and  $p$ .<sup>24</sup> Inspection of the expressions appearing in Theorem 1 reveals that the bounds become wider the larger the net bunching probability  $\mathcal{B} - p$ . When  $f_0(k) \approx f_1(k)$  and  $p = 0$ , the bounds will tend to be narrower when  $F_0(k)$  is closer to  $(1 - \mathcal{B})/2$ , i.e. the kink is close to the median of the latent hours distribution. This helps explain why the estimated bounds in Section 5 turn out to be quite informative.

*Comparison to existing results.* The existing bunching design literature does contain a few identification results that circle the common intuition that bunching is informative about a *local* average responsiveness, when responsiveness to incentives varies by observational unit. For instance, Saez (2010) and Kleven (2016) consider a “small-kink” approximation that  $\mathbb{E}[\Delta_{it} | h_{0it} = k] \approx \mathbb{B}/f_0(k)$ , where the RHS is observable and referred to as the “excess mass” quantity in Chetty

<sup>23</sup>It is worth noting that BLC of  $h_1$  and  $h_0$  implies bounds on the treatment effect  $Q_1(u) - Q_0(u)$  at *any* quantile  $u \in [0, 1]$ . But these bounds widen quickly as one moves away from the kink. The narrowest bounds for a single rank  $u$  are obtained for a “median” buncher roughly halfway between  $F_0(k)$  and  $F_1(k)$  when  $f_0(k) \approx f_1(k)$ . However, averaging over a larger group is more useful for meaningful ex-post evaluation of the FLSA, and reduces the sensitivity to departures from RANK (see Figure A.2). In the other extreme, one could drop RANK entirely and bound  $\mathbb{E}[h_{0it} - h_{it}]$  directly via BLC of  $h_0$  alone, but the bounds are *very* wide. The buncher ATE balances this tradeoff.

<sup>24</sup>Since the bounds depend only on the density around  $k$  and the total amount mass to its left/right, point masses elsewhere in the distributions of  $h_0$  and  $h_1$  have no effect on the bounds provided that they are well-separated from  $k$ .

et al., 2011.<sup>25</sup> The result requires  $f_0$  to be constant throughout the region  $[k, k + \Delta_{it}]$  conditional on each value of  $\Delta_{it}$ . This is most naturally justified by a kink that produces only tiny responses, an approximation that is likely to be quite poor when treatment corresponds to a 50% increase in the hourly cost of labor. Nevertheless, even in a small-kink setting Theorem 1 offers a refinement of this result. A second-order approximation to  $\ln(1 + \frac{x}{a})$  shows that when  $\mathcal{B}$  is small,  $\Delta_k^* \approx \frac{\mathcal{B}-p}{2f_0(k)} + \frac{\mathcal{B}-p}{2f_1(k)}$ . Another existing result comes from Blomquist et al. (2015): they show in a generic labor supply model that bunching identifies a certain weighted average of compensated elasticities, if the density of choices at a kink is assumed to be linear across counterfactual tax rates. But this parametric assumption is difficult to motivate, as these authors acknowledge.<sup>26</sup>

## 4.4 Estimating policy relevant parameters

The buncher ATE yields the answer to a particular causal question, among a well-defined subgroup of the population. Namely: how would hours among workers bunched at 40 hours by the overtime rule be affected by a counterfactual change from linear pay at their straight-time wages to linear pay at their overtime rates? This section discusses how we may now use this quantity to both evaluate the overall ex-post effect of the FLSA on hours, as well as forecast the impacts of proposed changes to the FLSA. This requires some additional assumptions, which I continue to approach from a partial identification perspective.

### 4.4.1 From the buncher ATE to the ex-post hours effect of the FLSA

To consider the overall ex-post hours effect of the FLSA among covered workers, I proceed in two steps. I first relate the buncher ATE to the overall average effect of introducing the overtime kink, holding fixed the distributions of counterfactual hours  $h_{0it}$  and  $h_{1it}$ . Then, I allow straight-time wages to be affected by the FLSA, using the buncher ATE again to bound the additional effect of these wage changes on hours.

To motivate this strategy, let us first define the parameter of interest to be the difference in average weekly hours with and without the FLSA:  $\theta := \mathbb{E}[h_{it}] - \mathbb{E}^*[h_{it}^*]$ , where  $h_{it}^*$  indicates the hours unit  $it$  would work absent the FLSA, and the second expectation  $\mathbb{E}^*$  is over units corresponding to workers that would exist in the no-FLSA counterfactual and be eligible were it introduced.<sup>27</sup> Defining  $\theta$  in this way allows us to remain agnostic as to whether the FLSA changes employment, and

---

<sup>25</sup>See Appendix A for a derivation in my generalized framework. The uniform density assumption is hard to justify except in the limit that the distribution of  $\Delta_{it}$  concentrates around zero. Proposition 6 in the Supplemental Material makes this claim precise, while connecting the approach from Saez (2010) and Kleven (2016) to results from Blomquist et al. (2015).

<sup>26</sup>In particular, the data identifies the density at the kink for two particular tax rates only (in the tax application), so cannot provide evidence of such linearity.

<sup>27</sup>Note that  $h_{it}^*$  in this section differs from the “anticipated” hours quantity  $h^*$  in Section 2.

hence the population of workers it applies to. However, I assume that the hours among any workers who enter or exit employment due to the FLSA are not systematically different from those who would exist anyways, so that we may rewrite  $\theta$  as  $\theta = \mathbb{E}[h_{it} - h_{it}^*]$ , an average over individual-level causal effects in the population that does exist given the FLSA.

Next, decompose  $\theta$  as:

$$\begin{aligned} \theta = \mathbb{E}[h_{it}(w_{it}, \mathbf{h}_{-i,t}) - h_{0it}(w_{it}^*, \mathbf{h}_{-i,t}^*)] &= \underbrace{\mathbb{E}[h_{it}(w_{it}, \mathbf{h}_{-i,t}) - h_{0it}(w_{it}, \mathbf{h}_{-i,t})]}_{\text{“effect of the kink”}} \\ &+ \underbrace{\mathbb{E}[h_{0it}(w_{it}, \mathbf{h}_{-i,t}) - h_{0it}(w_{it}^*, \mathbf{h}_{-i,t})]}_{\text{“wage effects”}} + \underbrace{\mathbb{E}[h_{0it}(w_{it}^*, \mathbf{h}_{-i,t}) - h_{0it}(w_{it}^*, \mathbf{h}_{-i,t}^*)]}_{\text{“interdependencies”}}, \end{aligned} \quad (8)$$

where the notation makes explicit the dependence of  $h$  and  $h_0$  on the worker’s straight-time wage  $w_{it}$ , and possibly the hours  $\mathbf{h}_{-i}$  of other workers in their firm. In the notation of the last section:  $h_{it} = h_{it}(w_{it}, \mathbf{h}_{-i,t})$ ,  $h_{0it} = h_{0it}(w_{it}, \mathbf{h}_{-i,t})$  and  $h_{1it} = h_{1it}(w_{it}, \mathbf{h}_{-i,t})$ . I have used that  $h_{it}^* = h_{0it}(w_{it}^*, \mathbf{h}_{-i,t}^*)$ , since pay is linear in hours in the no-FLSA counterfactual.

The first term in Equation (8) reflects the “effect of the kink” quantity  $h_{it} - h_{0it}$  examined in Section 4.2, and I view it as the first-order object of interest. The second term reflects that straight-time wages  $w_{it}$  may differ from those that workers would face without the FLSA, denoted by  $w_{it}^*$ . The third term is zero when firms’ choice of hours for its workers decomposes into separate optimization problems for each unit, as in the benchmark model from Section 4.2. More generally, it will capture any interdependencies in hours across units, for instance due to different workers’ hours being not linearly separable in production. In Appendix D I provide evidence that such effects do not play a large role in  $\theta$ , and I thus treat this term as zero when estimating  $\theta$ .<sup>28</sup>

Turning first to the “effect of the kink” term, note that with straight-wages and the hours of other units fixed, the kink only has such direct effects on those units working at least  $k = 40$  hours:

$$h_{it} - h_{0it} = \begin{cases} 0 & \text{if } h_{it} < k \\ k - h_{0it} & \text{if } h_{it} = k \\ -\Delta_{it} & \text{if } h_{it} > k \end{cases} \quad (9)$$

and thus  $\mathbb{E}[h_{it} - h_{0it}] = \mathcal{B} \cdot \mathbb{E}[k - h_{0it} | h_{it} = k] - P(h_{it} > k) \mathbb{E}[\Delta_{it} | h_{it} > k]$ . To identify this quantity we must extrapolate from the buncher ATE to obtain an estimate of  $\mathbb{E}[\Delta_{it} | h_{it} > k]$ , the average effect for units who work overtime. To do this, I assume that  $\Delta_{it}$  of units working more than

<sup>28</sup>In particular, I fail to find evidence of contemporaneous hours substitution in response to colleague sick pay, in an event study design. Another piece of evidence comes from obtaining similar “effect of the kink” estimates across small, medium and large firms, which suggests that a firm’s ability to reallocate hours between existing workers does not tend to drive their hours response to the FLSA. See Appendix D.



40 hours are at least as large on average as those who work 40, but that the reduced-form *elasticity* of their response is no greater than that of the bunchers. Assuming a constant percentage change between  $h_0$  and  $h_1$  over units would imply responses that grow in proportion to  $h_1$ , eventually becoming implausibly large. On the other hand, it would be an underestimate to assume high-hours workers, say at 60 hours, have the same effect in levels  $h_0 - h_1$  as those closer to 40.<sup>29</sup> Finally, to put bounds on the average effect of the kink among bunchers, I use bi-log-concavity of  $h_0$ . Details are provided in Supplemental Appendix 4.11.

The “wage effects” term in Equation (8) arises because the straight-time wages observed in the data may reflect some adjustment to the FLSA, as we would expect on the basis of the conceptual framework in Section 2. While the “effect of the kink” term is expected to be negative, this second term will be positive if FLSA causes a reduction in the straight-time wages set at hiring on the basis of expected hours. However, both terms ultimately depend on the same thing: responsiveness of hours to the cost of an hour of work. I thus use the buncher ATE to compute an approximate upper bound on wage effects by assuming that all straight-time wages are adjusted according to Equation (1) and that the hours response is iso-elastic in wages, with anticipated hours approximated by  $h_{it}$ . Supplemental Appendix 4.11 provides a visual depiction of these definitions. A lower bound on the “wage effects” term, on the other hand, is zero. Section 5 also reports results with and without this wage effect. The size of the wage effect  $\mathbb{E}[h_{0it} - h_{0it}^*]$  is appreciable but still small in comparison with  $\mathbb{E}[h_{it} - h_{0it}]$  (this is because the average percentage wage change according to Equation (1) is fairly small near 40, where most of the mass is).

#### 4.4.2 Forecasting the effects of policy changes

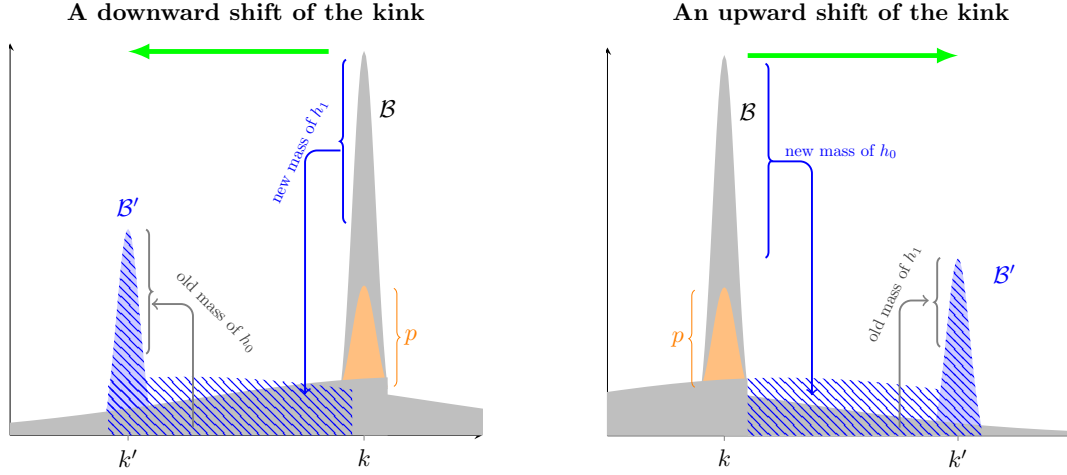
Apart from ex-post evaluation of the overtime rule, policymakers may also be interested in predicting what would happen if the parameters of overtime regulation were modified. Reforms that have been discussed in the U.S. include decreasing “standard hours”  $k$  at which overtime pay begins from 40 hours to 35 hours,<sup>30</sup> or increasing the overtime premium from time-and-a-half to “double-time” (Brown and Hamermesh, 2019). This section builds upon Sections 4.1 and 4.3 to show that the bunching-design model is also informative about the impact of such reforms on hours.

Let us begin by considering changes to standard hours  $k$ , for now holding the distributions of  $h_0$  and  $h_1$  fixed across the policy change. Inspection of Equation (2) reveals that as the kink is moved upwards, say from  $k = 40$  hours to  $k' = 44$  hours, some workers who were previously bunching at  $k$  now work  $h_{0it}$  hours: namely those for whom  $h_{0it} \in [k, k']$ . By the same token, some individuals with values of  $h_{1it} \in [k, k']$  now bunch at  $k'$ . Some individuals who were bunching at  $k$  now bunch

<sup>29</sup>In the model of Section 4.2, constant treatment effects in levels corresponds to exponential production:  $f(h) = \gamma(1 - e^{-h/\gamma})$  where  $\gamma > 0$  and  $h_{0it} - h_{1it} = \gamma \ln(1.5)$  for all units.

<sup>30</sup>Several countries have implemented changes to standard hours; Brown and Hamermesh (2019) provides a review of the evidence.

at  $k'$ —namely those for whom  $h_{1it} \leq k$  and  $h_{0it} \geq k'$ . In the case of a reduction in overtime hours, say to  $k' = 35$  this logic is reversed: some workers now work  $h_{1it} \in [k', k]$ , while workers with  $h_{0it} \in [k', k]$  now bunch at  $k'$ . I assume that the mass of counterfactual bunchers  $p$  remains at  $k = 40$  after the shift.<sup>31</sup> Figure 8 depicts both of these cases.



**FIGURE 7:** The left panel depicts a shift of the kink point downwards from  $k$  to  $k'$ , while right panel depicts a shift of the kink point upwards. See text for details.

Quantitatively assessing a change to double-time pay requires us to move beyond the two counterfactual choices  $h_{0it}$  and  $h_{1it}$ : hours that would be worked under straight-wage and time-and-a-half. Let  $h_{it}(\rho)$  be the hours that  $it$  would work if their employer faced a linear pay schedule at rate  $\rho \cdot w_{it}$  (with both the straight-wage  $w_{it}$  and hours of other units fixed at their realized levels). In this notation,  $h_{0it} = h_{it}(1)$  and  $h_{0it} = h_{it}(1.5)$ . Now consider a new overtime policy in which a premium pay factor of  $\rho_1$  is required for hours in excess of  $k$ , e.g.  $\rho_1 = 2$  for a “double-time” policy. Let  $h_{it}^{[k, \rho_1]}$  denote realized hours under this overtime policy, and let  $\mathcal{B}^{[k, \rho_1]} := P(h_{it}^{[k, \rho_1]} = k)$  the observable bunching that would occur.

Theorem 2 allows one to discuss the effects of small changes to  $k$  or  $\rho_1$ . Results for the effect of changing standard hours  $k$  (items 1. and 2. below) make use of an explicit assumption that firm preferences are quasi-linear with respect to costs:

**Assumption SEPARABLE.**  $\pi_{it}(z, \mathbf{x})$  is additively separable and linear in  $z$  for all units  $it$ .

I continue to assume that counterfactual bunchers  $K_{it}^* = 1$  stay at  $k^* := 40$ , regardless of  $\rho$  and  $k$ . To ease notation, let  $p(k) = p \cdot \mathbb{1}(k = k^*)$  denote the possible mass of counterfactual bunchers as a function of  $k$ .

<sup>31</sup>It is conceivable that some or all counterfactual bunchers locate at 40 because it is the FLSA threshold, while still being non-responsive to the incentives introduced there by the kink. In this case, we might imagine that they would all coordinate on  $k'$  after the change. The effects here should thus be seen as short-run effects before that occurs.

**Theorem 2 (marginal comparative statics in the bunching design).** *Under Assumptions CHOICE, CONVEX, SEPARABLE and SMOOTH:*

1.  $\partial_k \left\{ \mathcal{B}^{[k, \rho_1]} - p(k) \right\} = f_1(k) - f_0(k)$
2.  $\partial_k \mathbb{E}[h_{it}^{[k, \rho_1]}] = \mathcal{B}^{[k, \rho_1]} - p(k)$
3.  $\partial_{\rho_1} \mathcal{B}^{[k, \rho_1]} = -k f_{\rho_1}(k) \mathbb{E} \left[ \frac{dh_{it}(\rho_1)}{d\rho} \middle| h_{it}(\rho_1) = k \right]$
4.  $\partial_{\rho_1} \mathbb{E}[h_{it}^{[k, \rho_1]}] = - \int_k^\infty f_{\rho_1}(h) \mathbb{E} \left[ \frac{dh_{it}(\rho_1)}{d\rho} \middle| h_{it}(\rho_1) = h \right] dh$

*Proof.* See Appendix A. □

Assumption SMOOTH is a set of regularity conditions which imply that  $h_{it}(\rho)$  admits a density  $f_\rho(h)$  for all  $\rho$ ; see Appendix A for details. The above also uses a version of Assumption CHOICE given in Appendix A, which applies to all  $\rho$  rather than just  $\rho_0$  and  $\rho_1$ .

Beginning from the actual FLSA policy of  $k = 40$ ,  $\rho_1 = 1.5$ , the RHS of Items 1 and 2 are point identified from the data, provided that  $p$  is known. Item 1 says that if the location of the kink is changed marginally, the bunching probability will change according to the difference between the densities of  $h_{1i}$  and  $h_{0i}$  at  $k^*$ , which are in turn equal to the left and right limits of the observed density  $f(h)$  at the kink. This result is intuitive: given continuity of each potential outcome's density, a small increase in  $k$  will result in a mass proportional to  $f_1(k)$  being “swept in” to the mass point at the kink, while a mass proportional to  $f_0(k)$  is left behind. Item 2 aggregates this change in bunching with the changes to non-bunchers' hours as  $k$  is increased: the combined effect turns out to be to simply transport the mass of inframarginal bunchers to the new value of  $k$ .<sup>32</sup> Making use of Theorem 2 for a discrete policy change like reducing standard hours to 35 requires integrating across the actual range of hypothesized policy variation. We lose point identification, but I use bi-log-concavity of the marginal distributions of  $h_0$  and  $h_1$  to retain bounds.

Now consider the effect of moving from time-and-a-half to double time on average hours worked, in light of Item 4. This scenario, similar to the effect of the kink term in Eq. (8), requires making assumptions about the response of individuals who may locate far above the kink, and for whom the buncher ATE is less directly informative. First, we integrate Item 4 over  $\rho$  to obtain an expression for the average effect on hours from a move to double-time, which can be written in terms of local average elasticities of response:

$$\mathbb{E}[h_{it}^{[k, \rho_1]} - h_{it}^{[k, \bar{\rho}_1]}] = \int_{\rho_1}^{\bar{\rho}_1} d \ln \rho \int_k^\infty f_\rho(h) \cdot h \cdot \mathbb{E} \left[ \frac{d \ln h_{it}(\rho)}{d \ln \rho} \middle| h_{it}(\rho) = h \right] dh$$

---

<sup>32</sup>Intuitively, in the limit of a small change, bunchers who would choose exactly  $k$  under one of the two cost functions  $B_0$  or  $B_1$  cease to “bunch” as  $k$  increases, but they also do not change their realized  $h$ . Moore (2021) gives a closely-related result, derived independently of this work, that shows that bunching is a sufficient statistic for the effect of a marginal change in  $k$  on revenue in the context of a tax kink.

Recall that in the iso-elastic model the elasticity quantity  $\frac{d \ln h_{it}(\rho)}{d \ln \rho} = \frac{dh_{it}(\rho)}{d\rho} \frac{\rho}{h_{it}(\rho)}$  is constant across  $\rho$  and across units, and it is partially identified under BLC. I argue that just as an iso-elastic response is likely to overstate responsiveness at large values of hours, it is likely to *understate* responsiveness to larger values of  $\rho$ , thus yielding a lower bound on the effect of moving to double-time. For an upper bound on the magnitude of the effect, I assume rather that in levels  $\mathbb{E}[h_{it}(\rho_1) - h_{it}(\bar{\rho}_1) | h_{1it} > k]$  is at least as large as  $\mathbb{E}[h_{0it} - h_{1it} | h_{1it} > k]$ , and that the increase in bunching from a change of  $\rho_1$  to  $\bar{\rho}_1$  is as large as the increase from  $\rho_0$  to  $\rho_1$ . Additional details are provided in Supplemental Appendix 4.11.

In these calculations, I have held fixed the distributions of  $h_0$  and  $h_1$ . This can be seen as describing a short-run response before adjustment to straight-time wages or other factors that influence these latent hours distributions. In the empirical implementation I account for possible changes to straight wages when considering the average effects of policy changes on hours, as we saw with the ex-post effect of the FLSA.

## 5 Implementation and Results

This section implements the empirical strategy described in the last section with the sample of administrative payroll data described in Section 3.

### 5.1 Identifying counterfactual bunching at 40 hours

To deliver final estimates of the effect of the FLSA overtime rule on hours, it is necessary to first return to an issue raised in the introduction and alluded to in Section 4: that there are other reasons to expect bunching at 40 hours, in addition to being the location of the FLSA kink. For one, 40 may reflect a status-quo choice, being chosen even when it is not exactly profit maximizing for the firm. This effect could be amplified by firms synchronizing the schedules of different workers, requiring some common number of hours per week to coordinate around. Finally, for any salaried workers who were not successfully removed from the sample, firms might record the number of hours in a pay period as 40 even as actual hours worked vary.

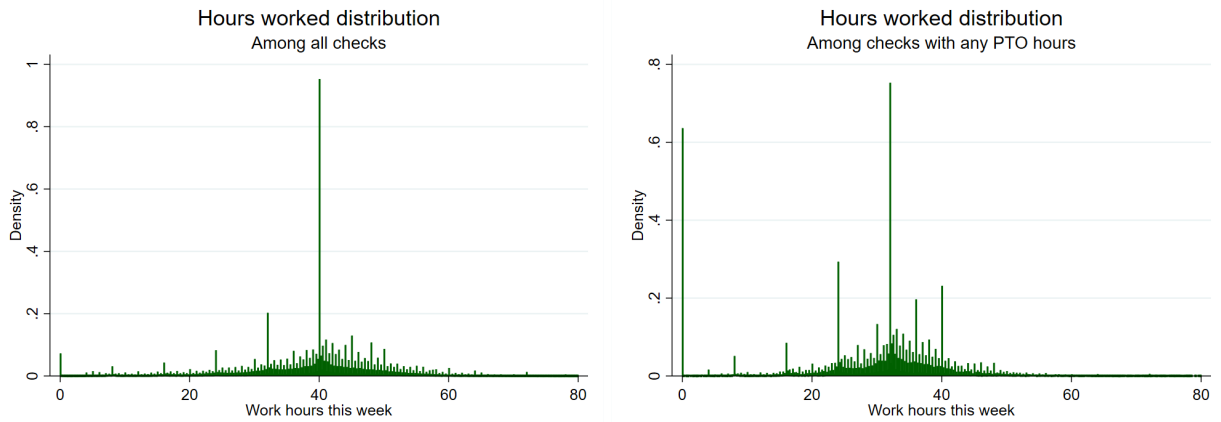
In terms of the empirical strategy from Section A.2, all of these alternative explanations manifest in the same way: a point mass  $p$  at 40 in the distribution of hours that would occur even if workers' pay did not feature a kink at 40. In the notation introduced in Section 4.3, these “counterfactual bunchers” are demarcated by  $K_{it}^* = 1$ ; I refer to the  $K_{it}^* = 0$  individuals who also locate at the kink as “active bunchers”. The mass of active bunchers is  $\mathcal{B} - p$ . Theorem 1 shows that we can still partially identify the buncher ATE in the presence of counterfactual bunchers, so long as we know what proportion of the total bunchers are active and how many are counterfactual.

I leverage two strategies to provide plausible estimates for the mass of counterfactual bunchers  $p$ . My preferred estimate uses of the fact that when an employee is paid for hours that are not actually worked—including sick time, paid time off (PTO) and holidays—these hours do not contribute to the 40 hour overtime threshold of the FLSA. For example, if a worker applies PTO to miss a six hour shift, then they are not required to be paid overtime until they reach 46 total paid hours in that week. Thus while the kink remains at 40 hours *worked*, non-work hours like PTO shift the location of the kink in hours of *pay*.

The identifying assumption that I rely on is that individuals who still work 40 hours a week, even when they take a positive number of non-work hours (and are hence paid for more than 40), are all active bunchers: they would not locate at forty hours in the counterfactuals  $h_{0it}$  and  $h_{1it}$ . This reflects the idea that additional explanations for bunching at 40 hours operate at the level of hours paid, rather than hours worked. Let  $n_{it}$  indicate non-work hours of pay for worker  $i$  in week  $t$ . Specifically, I make the following two assumptions:

1.  $P(h_{it} = 40 | n_{it} > 0) = P(h_{it} = 40 \text{ and } K_{it}^* = 0 | n_{it} > 0)$
2.  $P(h_{it} = 40 \text{ and } K_{it}^* = 0 | n_{it} > 0) = P(h_{it} = 40 \text{ and } K_{it}^* = 0 | n_{it} = 0)$

The first item allows me to identify the mass of active bunchers in the  $n_{it} > 0$  conditional distribution of hours. The second item says that this conditional mass is representative of the unconditional mass of active bunchers. To increase the plausibility of this assumption, I focus on  $\eta$  as paid time off because it is generally planned in advance, yet has somewhat idiosyncratic timing. By contrast, sick pay is often unanticipated so the firm may not be able to re-optimize total hours within the week in which a worker calls in sick. Holiday pay is known in advance, but holidays are unlikely to be representative in terms of other factors important for hours determination (e.g. product demand).



**FIGURE 8:** The right panel shows a histogram of hours worked when paid time off hours are positive ( $\eta_{it} > 0$ ). The left panel shows the unconditional distribution. While  $\mathcal{B} \approx 11.6\%$ ,  $P(h_{it} = 40 | n_{it} > 0) \approx 2.7\%$ .

Together, the two assumptions above imply that  $p = P(K_{it}^* = 1 \text{ and } h_{it} = 40)$  is identified as  $\mathcal{B} - P(h_{it} = 40 | \eta_{it} > 0)$ . Figure 8 shows the conditional distribution of hours paid for work when the paycheck contains a positive number of PTO hours ( $n_{it} > 0$ ). The figure reveals that when moving from the unconditional (left panel) to positive-PTO conditional (right panel) distribution, most of the point mass at 40 hours moves away, largely concentrating now at 32 hours (corresponding to the PTO covering a single eight hour shift). Of the total bunching of  $\mathcal{B} \approx 11.6\%$  in the unconditional distribution, I estimate that only about  $P(h_{it} = 40 | n_{it} > 0) \approx 2.7\%$  are active bunchers, leaving  $p \approx 8.9\%$ . Thus roughly three quarters of the individuals at 40 hours are counterfactual rather than active bunchers.

As a secondary strategy, I estimate an upper bound for  $p$  by using the assumption that the potential outcomes of counterfactual bunchers are relatively “sticky” over time. If the hours of counterfactual bunchers are at forty for behavioral or administrative reasons, it may be reasonable to assume that these external considerations are fairly static, preventing latent hours  $h_{0it}$  from changing much between adjacent weeks. In particular, assume that in a given week  $t$  nearly all of the counterfactual bunchers are also non-movers from week  $t - 1$ . Then:

$$p = P(h_{0it} = 40) \approx P(h_{0it} = h_{0it-1} = 40) \leq P(h_{it} = h_{i,t-1} = 40),$$

where the inequality follows from  $(h_{0it} = 40) \implies (h_{it} = 40)$  by Lemma 1. The probability  $P(h_{it} = h_{i,t-1} = 40)$  can be directly estimated from the data, yielding  $p \leq 6\%$ .

## 5.2 Estimation and inference

Given Theorem 1 and a value of  $p$ , computing bounds on the buncher ATE requires estimates of the right and left limits of the CDF and density of hours at the kink. I use the local polynomial density estimator of Cattaneo, Jansson and Ma (2020) (CJM), which is well-suited to estimating a CDF and its derivatives at boundary points. For instance, a local-linear CJM estimator provides a smoothed estimate of the left limit of the CDF and density at  $k$  as:

$$(\hat{F}_-(k), \hat{f}_-(k)) = \underset{(b_1, b_2)}{\operatorname{argmin}} \sum_{it: h_{it} < k} (F_n(h_{it}) - b_1 - b_2 h_{it})^2 \cdot K\left(\frac{h_{it} - k}{\alpha}\right) \quad (10)$$

where  $F_n(y) = \frac{1}{n} \sum_{it} \mathbb{1}(h_{it} \leq y)$  is the empirical CDF of a sample of size  $n$ ,  $K(\cdot)$  is a kernel function, and  $\alpha$  is a bandwidth. The right limits  $F_+(k)$  and  $f_+(k)$  are estimated analogously using observations for which  $h_{it} > k$ . I use a triangular kernel, and choose  $h$  as follows: first, I use CJM’s mean-squared error minimizing bandwidth selector to produce a bandwidth choice using the data on either side of  $k = 40$  (for the left and right limits, respectively). I then average the two bandwidths, and use this as the bandwidth in the final calculation of both the right and left limits,

to mitigate any dependence on a differential bandwidth choice for each side. In the full sample, the bandwidth chosen by this procedure is about 1.7 hours, and is somewhat larger for subsamples that condition on a single industry.

To construct confidence intervals for parameters that are partially identified (e.g. the buncher ATE), I use the adaptive critical values proposed by Imbens and Manski (2004) and Stoye (2009) that are valid for the underlying parameter. In each case, estimators of bounds or point identified quantities are functions of inputs that are  $\sqrt{n}$ -asymptotically normal. To easily incorporate sampling uncertainty in  $\hat{F}_-(k)$ ,  $\hat{f}_-(k)$ ,  $\hat{F}_+(k)$ ,  $\hat{f}_+(k)$  and  $\hat{p}$ , I estimate the variances by a cluster nonparametric bootstrap that resamples at the firm level. This allows arbitrary autocorrelation in hours across pay periods for a single worker, and between workers within a firm. All standard errors use 500 bootstrap replications.

### 5.3 Results of the bunching estimator: the buncher ATE

Table 3 reports treatment effect estimates based on Theorem 1, when  $p$  is either assumed zero or estimated by one of the two methods described in Section 5.1. These estimates use a sample that pools across workers in all industries. The first row reports the corresponding estimate of the net bunching probability  $\mathcal{B} - p$ , while the second row reports the bounds on the buncher ATE  $\mathbb{E}[h_{0it} - h_{1it} | h_{it} = k]$ . Within a fixed estimate of  $p$ , the bounds on the buncher ATE based on bi-log-concavity are quite informative: the upper and lower bounds are always close to each other and precisely estimated. Appendix F reports estimates based on alternative shape constraints and assumptions about effect heterogeneity, which deliver similar results.<sup>33</sup>

The PTO-based estimate of  $p$  provides the most conservative treatment effect estimates, attributing roughly one quarter of the observed bunching to active rather than counterfactual bunchers. Nevertheless, this estimate still yields a highly statistically significant buncher ATE of about 2/3 of an hour, or 40 minutes. This estimate can be interpreted as follows: consider the group of workers that are in fact working 40 hours in a given pay period and are not counterfactual bunchers. This group would work on average about 40 minutes more that week if they were paid their straight-time wage for all hours, compared with a counterfactual in which they are paid their overtime rate for all hours. If we instead attribute all of the observed bunching mass to active bunchers ( $p = 0$ ), then this buncher ATE parameter is estimated to be at least 2.6 hours. In Appendix F I report estimates of the buncher ATE for each of the largest industries in the sample, and also plot estimates directly as a function of the assumed mass  $p$  of counterfactual bunchers at 40 hours.

<sup>33</sup>In particular, I present a point estimate based on Appendix Proposition 1, which assumes that treatment effects are constant and that the density is linear in the missing region, as well as results under a weaker assumption that the density is monotonic in the missing region. Monotonicity is not likely to hold in the overtime context, but the bounds based on monotonicity do not deliver vastly different results. See Tables F.10 and F.11, which applies the same assumptions to the distribution of log hours rather than hours.

	$p=0$	$p$ from non-changers	$p$ from PTO
Net bunching:	0.116 [0.112, 0.120]	0.057 [0.055, 0.058]	0.027 [0.024, 0.030]
Buncher ATE	[2.614, 3.054] [2.493, 3.205]	[1.324, 1.435] [1.264, 1.501]	[0.640, 0.666] [0.574, 0.736]
Num observations	630217	630217	630217
Num clusters	566	566	566

**TABLE 3:** Estimates of net bunching  $\mathcal{B} - p$  and the buncher ATE:  $\Delta_k^* = \mathbb{E}[h_{0it} - h_{1it} | h_{it} = k, K_{it}^* = 0]$ , across various strategies to estimate counterfactual bunching  $p = P(K_{it}^* = 1)$ . Unit of analysis is a paycheck, and 95% bootstrap confidence intervals (in gray) are clustered by firm.

## 5.4 Estimates of policy effects

I now use estimates of the buncher ATE and the results of Section 4.4 to estimate the overall causal effect of the FLSA overtime rule, and simulate changes based on modifying standard hours or the premium pay factor. Table 4 first reports an estimate of the buncher ATE expressed as a reduced-form hours demand elasticity,<sup>34</sup> which I use as an input in these calculations. The next two rows report bounds on  $\mathbb{E}[h_{it} - h_{it}^*]$  and  $\mathbb{E}[h_{it} - h_{it}^* | h_{1it} \geq 40, K_{it}^* = 0]$ , respectively. The second row is the overall ex-post effect of the FLSA on hours, averaged over workers and pay periods, and the third row conditions on paychecks reporting at least 40 hours (omitting counterfactual bunchers). The final row reports an estimate of the effect of moving to double-time pay. I provide details of the calculations in Supplemental Appendix 4.11.

Taking the PTO-based estimate of  $p$  as yielding a lower bound on treatment effects, the estimates suggest that FLSA eligible workers work at least 1/5 of an hour less in any given week than they would absent overtime regulation: about one third the magnitude of the buncher ATE in levels. When I focus on those eligible workers that are directly affected in a given week, the figure is about twice as high: roughly 30 minutes. I estimate that a move to double-time pay would introduce a further reduction that may be comparable to the existing overall ex-post effect, but with substantially wider bounds. These estimates include the effects of possible adjustments to straight-time wages, which tend to attenuate the effects of the policy change. Appendix Table F.12 replicates

<sup>34</sup> This is  $\hat{\Delta}_k^* / (40 \ln(1.5))$  where  $\hat{\Delta}_k^*$  is the estimate of the buncher ATE presented in Table 3, which is numerically equivalent to the elasticity implied by the buncher ATE in logs  $\mathbb{E}[\ln h_{0it} - \ln h_{1it} | h_{it} = k, K_{it}^* = 0] / (\ln 1.5)$  estimated under assumption that  $\ln h_0$  and  $\ln h_1$  are BLC.



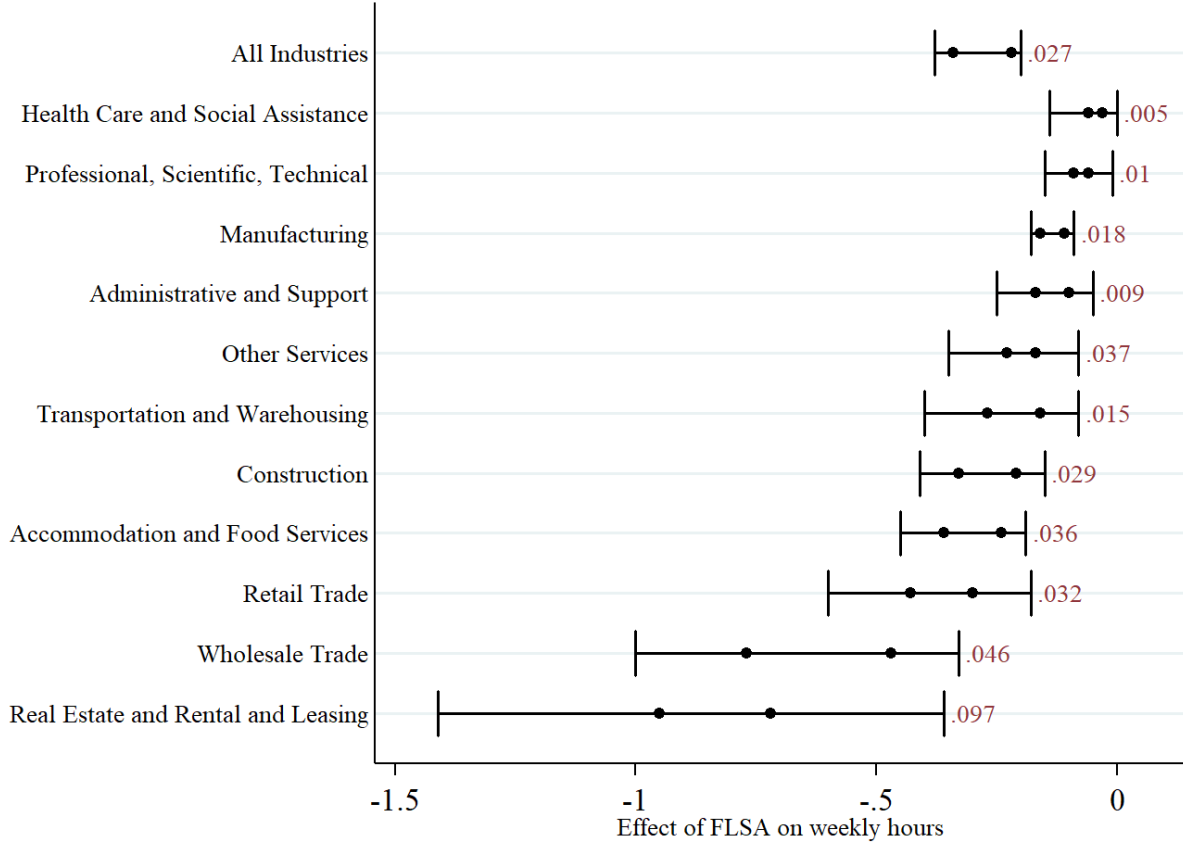
Table 4 neglecting these wage adjustments, which might be viewed as a short-run response to the FLSA before wages have time to adjust.

	$p=0$	$p$ from non-changers	$p$ from PTO
Buncher ATE as elasticity	[-0.188,-0.161] [-0.198,-0.154]	[-0.088,-0.082] [-0.093,-0.078]	[-0.041,-0.039] [-0.045,-0.035]
Average effect of FLSA on hours	[-1.466, -1.026] [-1.535, -0.977]	[-0.727, -0.486] [-0.762, -0.463]	[-0.347, -0.227] [-0.384, -0.203]
Avg. effect among directly affected	[-2.620, -1.833] [-2.733, -1.750]	[-1.453, -0.972] [-1.518, -0.929]	[-0.738, -0.483] [-0.812, -0.434]
Double-time, average effect on hours	[-2.604, -0.569] [-2.707, -0.547]	[-1.239, -0.314] [-1.285, -0.300]	[-0.580, -0.159] [-0.638, -0.143]

**TABLE 4:** Estimates of the buncher ATE expressed as an elasticity, the average ex-post effect of the FLSA  $\mathbb{E}[h_{it} - h_{it}^*]$ ,<sup>34</sup> the effect among directly affected units  $\mathbb{E}[h_{it} - h_{it}^* | h_{it} \geq k]$  and predicted effects of a change to double-time. 95% bootstrap confidence intervals in gray, clustered by firm.

Figure 9 breaks down estimates of the ex-post effect of the overtime rule by major industries, revealing considerable heterogeneity between them. The estimates suggest that Real Estate & Rental and Leasing as well as Wholesale Trade see the highest average reduction in hours. The least-affected industries are Health Care and Social Assistance and Professional Scientific and Technical, with the average worker working just about 6 minutes less per week. Appendix Figure F.11 compares the hours distribution for Real Estate & Rental and Leasing with the distribution for of Professional Scientific and Technical, showing that the difference in their effects is explained both by a larger value of  $\mathcal{B} - p$  and a lower density of hours close to the kink for Real Estate & Rental and Leasing. Appendix Table F.6 reports the numerical estimates and confidence intervals by industry. Appendix F reports estimates broken down by gender, finding that the FLSA has considerably higher effects on the hours of men compared with women.

Appendix Figure F.9 looks at the effect of changing the threshold for overtime hours  $k$  from 40 to alternative values  $k'$ . The left panel reports estimates of the identified bounds on  $\mathcal{B}^{[k', \rho_1]}$  as well as point-wise 95% confidence intervals (gray) across values of  $k'$  between 35 and 45, for each of the three approaches to estimating  $p$ . In all cases, the upper bound on bunching approaches zero as  $k'$  is moved farther from 40. This is sensible if the  $h_0$  and  $h_1$  distributions are roughly unimodal with modes around 40: straddling of potential outcomes becomes less and less likely as one moves away from where most of the mass is. Appendix F.10 shows these bounds as  $k'$  ranges all the way from 0 to 80, for the  $p = 0$  case. These estimates should be viewed as short-run responses, as they



**FIGURE 9:** 95% confidence intervals for the effect of the FLSA on hours by industry, using PTO-based estimates of  $p$  for each. Dots are point estimates of the upper and lower bounds. The number to the right of each range is the point estimate of the net bunching  $B - p$  for that industry.

do not account for adjustment to straight-time wages.

When  $p$  is estimated using PTO or non-changers between periods, we see that the upper bound of the identified set for  $\mathcal{B}^{[k', \rho_1]}$  in fact reaches zero quite quickly in  $k'$ . Moving standard hours to 35 is predicted to completely eliminate bunching due to the overtime kink in the short run, before any adjustment to latent hours (e.g. through changes to straight-time wages). The right panel of Appendix Figure F.9 shows estimates for the average effect on hours of changing standard hours, inclusive of wage effects (see Appendix 4.11 for details). Increases to standard hours cause an increase in hours per worker, as overtime policy becomes less stringent, and reductions to standard hours reduce hours. The magnitude of these effects is not precisely estimated for changes larger than a couple of hours, however the range of statistically significant effects depends on  $p$ . Even for the preferred estimate of  $p$  from PTO, increasing the overtime threshold as high as 43 hours is estimated to increase average working hours by an amount distinguishable from zero.

## 6 Implications of the estimates for overtime policy

The estimates from the preceding section suggest that FLSA regulation indeed has real effects on hours worked, in line with labor demand theory when wages do not fully adjust to absorb the added cost of overtime hours. When averaged over affected workers and across pay periods, I find that hourly workers in my sample work at least 30 minutes less per week than they would without the overtime rule. This lower bound is broadly comparable to the few causal estimates that exist in the literature, including Hamermesh and Trejo (2003) who assess the effects of expanding California’s daily overtime rule to cover men in 1980, and Brown and Hamermesh (2019) who use the erosion of the real value of FLSA exemption thresholds over the last several decades.<sup>35</sup> By contrast, my estimates carry the strengths of an approach to identification that does not require focusing on the sub-population affected by a natural experiment, and use much more recent data.

These estimates speak to the substitutability of hours of labor between workers. The primary justifications for overtime regulation have been to reduce excessive workweeks, while encouraging hours to be distributed over more workers (Ehrenberg and Schumann, 1982). How well this—and related policies such as work-sharing programs—play out in practice hinges on how easily an hour of work can be moved from one worker to another or across time, from the perspective of the firm. The results of this paper find hours demand to be relatively inelastic: hours cannot be easily so reallocated between workers or weeks. This suggests that ongoing efforts to expand coverage of the FLSA overtime rule (by increasing the earnings threshold at which some salaried workers are exempt) may have limited scope to drastically affect the hours of U.S. workers.

Nevertheless, the overall impact of the FLSA overtime rule on workers could be substantial. The data suggest that at least about 3% and as many as about 12% of workers’ hours are adjusted to the threshold introduced by the policy, indicating that it may have distortionary impacts for a significant portion of the labor force. The policy may also have important effects on unemployment. While a full assessment of the employment effects of the FLSA overtime rule is beyond the scope of this paper, my estimates of the hours effect can be used to build a back-of-the-envelope calculation. Following Hamermesh (1996), I assume a value for the rate at which firms substitute labor for capital to obtain a “best-guess” estimate that the FLSA overtime rule creates about 700,000 jobs (see Appendix F.6 for details). To get an overall upper bound on the size of employment effects, I attribute all of the bunching at 40 to the FLSA and assume that the total number of worker-hours is not reduced by the FLSA. By this estimate the FLSA increases employment by at most 3 million jobs, or 3% among covered workers. A reasonable range of parameter values rules out negative

---

<sup>35</sup>Hamermesh and Trejo (2003) and Brown and Hamermesh (2019) report estimates of  $-0.5$  and  $-0.18$  for the elasticity of overtime hours with respect to the overtime rate. My preferred estimate of  $-0.04$  for the buncher ATE as an elasticity is the elasticity of *total* hours, including the first 40. An elasticity of overtime hours can be computed from this using the ratio of mean hours to mean overtime hours in the sample, resulting in an estimate of roughly  $-0.45$ .

overall employment effects from the FLSA.

I have also considered the likely effects of adjusting the two parameters that characterize the FLSA overtime rule: standard hours and the overtime premium factor. Estimates suggest that moving to double-pay for overtime would have an average additional effect on hours that is at least as large as the effect of the current FLSA regulation, and moving standard hours to 35 rather than 40 would nearly eliminate bunching due to the FLSA, given workers' current wages.<sup>3637</sup>

## 7 Conclusion

This paper has provided a reinterpretation of the bunching-design method in the language of treatment effects, showing that the basic identifying power of the method is robust to a variety of underlying structural choice models. Across such modeling choices, the parameter of interest remains a reduced-form local average treatment effect between two appropriately-defined counterfactual choices, which can be partially identified by extrapolating nonparametric shape constraints on counterfactual distributions. This provides conditions under which the bunching design can be useful to answer program evaluation questions in a broad variety of contexts, particularly beyond those in which the researcher is prepared to posit a parametric model of decision-makers' preferences.

By leveraging these insights with a new payroll dataset recording exact weekly hours paid at the individual level, I estimate that U.S. workers subject to the Fair Labor Standard Act work shorter hours due to its overtime provision, which may lead to positive employment effects. Given the large amount of within-worker variation in hours observed in the data, the modest size of the FLSA effects estimated in this paper suggest that firms do face significant incentives to maintain longer working hours, countervailing against the ones introduced by policies intended to reduce them.

## References

- BARKUME, A. (2010). "The Structure of Labor Costs with Overtime Work in U.S. Jobs". *Industrial and Labor Relations Review* 64 (1).
- BERTANHA, M., MCCALLUM, A. H. and SEEGER, N. (2020). "Better Bunching , Nicer Notching". *SSRN Working Paper*.

---

<sup>36</sup>Estimates of the average hours effect for changes to standard hours are consistent with estimates by Costa (2000), that hours fell by 0.2-0.4 on average during the phased introduction of the FLSA in which standard hours declined by 2 hours in 1939 and 1940.

<sup>37</sup>While my short run prediction under this policy counterfactual assumes away changes to straight-time wages, the reduction in bunching is likely to remain after allowing such adjustment over time. With 35 already to the left of the mode of the latent hours distributions  $h_{0i}$  and  $h_{1i}$ , it would become even further from the mode as these distributions move rightward due to lower wages.

- BEST, M. C., BROCKMEYER, A., KLEVEN, H. J., SPINNEWIJN, J. and WASEEM, M. (2015). "Production vs Revenue Efficiency With Limited Tax Capacity: Theory and Evidence From Pakistan". *Journal of Political Economy* 123 (6), p. 48.
- BISHOW, J. L. (2009). "A Look at Supplemental Pay: Overtime Pay, Bonuses, and Shift Differentials". *Monthly Labor Review*. Publisher: Bureau of Labor Statistics, U.S. Department of Labor.
- BLOMQUIST, S., KUMAR, A., LIANG, C.-Y. and NEWHEY, W. (2021). "On Bunching and Identification of the Taxable Income Elasticity". *Journal of Political Economy* Forthcoming.
- BLOMQUIST, S., KUMAR, A., LIANG, C.-Y. and NEWHEY, W. K. (2015). "Individual heterogeneity, nonlinear budget sets and taxable income". *The Institute for Fiscal Studies Working Paper* CWP21/15.
- BLOMQUIST, S. and NEWHEY, W. (2017). "The Bunching Estimator Cannot Identify the Taxable Income Elasticity". *The Institute for Fiscal Studies Working Paper* CWP40/17.
- BRECHLING, F. P. R. (1965). "The Relationship Between Output and Employment in British Manufacturing Industries". *The Review of Economic Studies* 32 (3), p. 187.
- BROWN, C. and HAMERMESH, D. S. (2019). "Wages and Hours Laws: what do we know? what can be done?" *The Russell Sage Foundation Journal of the Social Sciences* 5 (5), pp. 68–87.
- BURDETT, K. and MORTENSEN, D. T. (1998). "Wage Differentials, Employer Size, and Unemployment". *International Economic Review* 39 (2), p. 257.
- CAETANO, C., CAETANO, G. and NIELSEN, E. (2020). "Correcting for Endogeneity in Models with Bunching". *Federal Reserve Board Finance and Economics Discussion Series* 2020-080.
- CAHUC, P. and ZYLBERBERG, A. (2004). *Labor economics*. OCLC: 265445233. Cambridge, Mass.: MIT Press.
- CATTANEO, M. D., JANSSON, M. and MA, X. (2020). "Simple Local Polynomial Density Estimators". *Journal of the American Statistical Association* 115 (531), pp. 1449–1455.
- CHERNOZHUKOV, V. and HANSEN, C. (2005). "An IV Model of Quantile Treatment Effects". *Econometrica* 73 (1), pp. 245–261.
- CHETTY, R., FRIEDMAN, J. N., OLSEN, T. and PISTAFERRI, L. (2011). "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126 (2), pp. 749–804.

- COSTA, D. L. (2000). “Hours of Work and the Fair Labor Standards Act: A Study of Retail and Wholesale Trade, 1938–1950”. *Industrial and Labor Relations Review*, p. 17.
- DÜMBGEN, L., KOLESNYK, P. and WILKE, R. A. (2017). “Bi-log-concave distribution functions”. *Journal of Statistical Planning and Inference* 184, pp. 1–17.
- DUBE, A., MANNING, A. and NAIDU, S. (2020). “Monopsony, Misoptimization, and Round Number Bunching in the Wage Distribution”. *NBER Working Paper* w24991.
- EHRENBERG, R. and SCHUMANN, P. (1982). *Longer hours or more jobs? : an investigation of amending hours legislation to create employment*. New York State School of Industrial and Labor Relations, Cornell University.
- EHRENBERG, R. G. (1971). “The Impact of the Overtime Premium on Employment and Hours in U . S . Industry”. *Economic Inquiry* 9 (2).
- EINAV, L., FINKELSTEIN, A. and SCHRIMPF, P. (2017). “Bunching at the kink: Implications for spending responses to health insurance contracts”. *Journal of Public Economics* 146, pp. 27–40.
- GELBER, A. M., JONES, D. and SACKS, D. W. (2020). “Estimating Adjustment Frictions Using Nonlinear Budget Sets: Method and Evidence from the Earnings Test”. *American Economic Journal: Applied Economics* 12 (1), pp. 1–31.
- GRIGSBY, J., HURST, E. and YILDIRMAZ, A. (2020). *Aggregate Nominal Wage Adjustments: New Evidence from Administrative Payroll Data*. *American Economic Review*, forthcoming.
- HAMERMESH, D. S. (1996). *Labor demand*. Princeton, NJ: Princeton Univ. Press.
- HAMERMESH, D. S. and TREJO, S. J. (2003). “The Demand for Hours of Labor : Direct Evidence from California”. *The Review of Economics and Statistics* 82 (1), pp. 38–47.
- HART, R. A. (2004). *The economics of overtime working*. OCLC: 704550114. Cambridge, UK: Cambridge University Press.
- HJORT, J., LI, X. and SARSONS, H. (2020). “Across-Country Wage Compression in Multinationals”. *NBER Working Paper* w26788.
- HUANG, C.-i. (2008). “Estimating demand for cellular phone service under nonlinear pricing”. *Quantitative Marketing and Economics volume* 6, pp. 371–413.

- IMBENS, G. W. and MANSKI, C. F. (2004). “Confidence Intervals for Partially Identified Parameters”. *Econometrica* 72, p. 14.
- ITO, K. and SALLEE, J. M. (2017). “The Economics of Attribute-Based Regulation: Theory and Evidence from Fuel-Economy Standards”. *The Review of Economics and Statistics*, pp. 319–336.
- JOHNSON, J. (2003). “The Impact of Federal Overtime Legislation on Public Sector Labor Markets”. *Journal of Labor Economics* 21 (1), pp. 43–69.
- KASY, M. (2017). “Who wins, who loses? Identification of the welfare impact of changing wages”. *Working Paper*, pp. 1–26.
- KLEVEN, H. J. (2016). “Bunching”. *Annual Review of Economics* 8 (June), pp. 435–464.
- KLEVEN, H. J. and WASEEM, M (2013). “Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan”. *The Quarterly Journal of Economics* 128 (2), pp. 669–723.
- LEWIS, H. G. (1969). “Employer Interest in Employee Hours of Work”. *Unpublished paper*.
- MILGROM, P. and ROBERTS, J. (1996). “The LeChatelier Principle”. *American Economic Review* 1 (86), pp. 173–179.
- MOORE, D. T. (2021). “Evaluating Tax Reforms without Elasticities: What Bunching Can Identify”, p. 61.
- QUACH, S. (2021). “The Labor Market Effects of Expanding Overtime Coverage”. *MPRA Paper* 100613.
- ROSEN, S. (1968). “Short-Run Employment Variation on Class-I Railroads in the U.S., 1947-1963”. *Econometrica* 36 (3), p. 511.
- SAEZ, E. (2010). “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy* 2 (3). ISBN: 1945-7731 \_eprint: arXiv:1011.1669v3, pp. 180–212.
- SOCIETY FOR HUMAN RESOURCE MANAGEMENT (2018). “National Study of Employers”, p. 79.
- STOLE, L. A. and ZWIEBEL, J. (1996). “Intra-Firm Bargaining under Non-Binding Contracts”. *The Review of Economic Studies* 63 (3). Publisher: [Oxford University Press, Review of Economic Studies, Ltd.], pp. 375–410.

- STOYE, J. (2009). “More on Confidence Intervals for Partially Identified Parameters”. *Econometrica* 77 (4), pp. 1299–1315.
- TREJO, B. S. J. (1991). “The Effects of Overtime Pay Regulation on Worker Compensation”. *American Economic Review* 81 (4), pp. 719–740.
- U.S. DEPARTMENT OF LABOR (2019). “Defining and Delimiting the Exemptions for Executive, Administrative, Professional, Outside Sales and Computer Employees”. *Federal Register* 84 (188).

## A Identification in a generalized bunching design

This section present a generalization of the bunching-design model used in the paper. While the FLSA will provide a running example throughout, I largely abstract from the overtime context to emphasize the general applicability of the results.

To facilitate comparison with the existing literature on bunching at kinks – which has mostly considered cross-sectional data – I throughout this section suppress time indices and use the single index  $i$  to refer to each unit of observation (a paycheck in the overtime case). Further, the “running variable” of the bunching design is denoted throughout this section by  $Y$  rather than  $h$ . This is done to emphasize the link to the treatment effects literature, while allowing a distinction that is in some cases important (e.g. a model where hours of pay for work differ from actual hours of work).

### A.1 A generalized bunching-design model

Consider a population of observational units indexed by  $i$ . For each  $i$ , a decision-maker  $d(i)$  chooses a point  $(z, \mathbf{x})$  in some space  $\mathcal{X} \subseteq \mathbb{R}^{m+1}$  where  $z$  is a scalar and  $\mathbf{x}$  a vector of  $m$  components, subject to a constraint of the form:

$$z \geq \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\} \quad (\text{A.1})$$

We require that  $B_{0i}(\mathbf{x})$  and  $B_{1i}(\mathbf{x})$  are continuous and weakly convex functions of the vector  $\mathbf{x}$ , and that there exist continuous scalar functions  $y_i(\mathbf{x})$  and a scalar  $k$  such that:

$$B_{0i}(\mathbf{x}) > B_{1i}(\mathbf{x}) \text{ whenever } y_i(\mathbf{x}) < k \quad \text{and} \quad B_{0i}(\mathbf{x}) < B_{1i}(\mathbf{x}) \text{ whenever } y_i(\mathbf{x}) > k$$



The value  $k$  is taken to be common to all units  $i$ , and is assumed to be known by the researcher.<sup>38</sup> In the overtime setting,  $y_i(\mathbf{x})$  represents the hours of work for which a worker is paid in a given week, and  $k = 40$ . In most applications of the bunching design, the decision-maker  $d(i)$  is simply  $i$  themselves, for example a worker choosing their labor supply subject to a tax kink. In the overtime application however  $i$  is a worker-week pair, and  $d(i)$  is that worker's firm.

Let  $X_i$  be  $i$ 's realized outcome of  $\mathbf{x}$ , and  $Y_i = y_i(X_i)$ . I assume that  $Y_i$  is observed by the econometrician, but not that  $X_i$  is.

In a typical example, the functions  $B_{0i}$ ,  $B_{1i}$  will represent a schedule of some kind of “cost” as a function of the choice vector  $\mathbf{x}$ , with two regimes of costs that are separated by the condition  $y_i(\mathbf{x}) = k$ , characterizing the locus of points at which the two cost functions cross. Let  $B_{ki}(\mathbf{x}) := \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\}$ . Budget constraints like Eq.  $z \geq B_{ki}(\mathbf{x})$  are typically “kinked” because while the function  $B_{ki}(\mathbf{x})$  is continuous, it will generally be non-differentiable at the  $\mathbf{x}$  for which  $y_i(\mathbf{x}) = k$ .<sup>39</sup> While the functions  $B_0$ ,  $B_1$  and  $y$  can all depend on  $i$ , I will often suppress this dependency for clarity of notation.

In the most common cases from the literature,  $\mathbf{x}$  is assumed to be the scalar  $y_i(x) = x$ , i.e. there is no distinction between the “kink variable”  $y$  and underlying choice variables  $\mathbf{x}$ . For example, the seminal bunching design papers Saez (2010) and Chetty et al. (2011) considered progressive taxation with  $z$  being tax liability (or credits), both  $y = x$  corresponding to taxable income, and  $B_0$  and  $B_1$  linear tax functions on either side of a threshold  $y$  between two adjacent tax/benefit brackets. However, even when the functions  $B_0$  and  $B_1$  only depend on  $\mathbf{x}$  through  $y_i(\mathbf{x})$ , the bunching design is compatible with models in which multiple margins of choice respond to the incentives provided by the kink.<sup>40</sup> In fact, the econometrician may be agnostic as to even what the full set of components of  $\mathbf{x}$  are, with  $y(\cdot)$ ,  $B_0(\cdot)$  or  $B_1(\cdot)$  depending only on various subsets of them. The next section will discuss how the bunching design allows us to conduct causal inference on the variable  $Y_i$ , but not directly on the underlying choice variables  $X_i$ .

In the overtime context,  $z$  corresponds to the cost of a single-worker's labor in a single week, and in the simplest models of hours choice the vector  $\mathbf{x}$  is simply equal to hours of work  $y$ . The

---

<sup>38</sup>This comes at little cost of generality since with heterogeneous  $k_i$  this could be subsumed as a constant into the function  $y_i(\mathbf{x})$ , so long as the  $k_i$  are observed by the researcher.

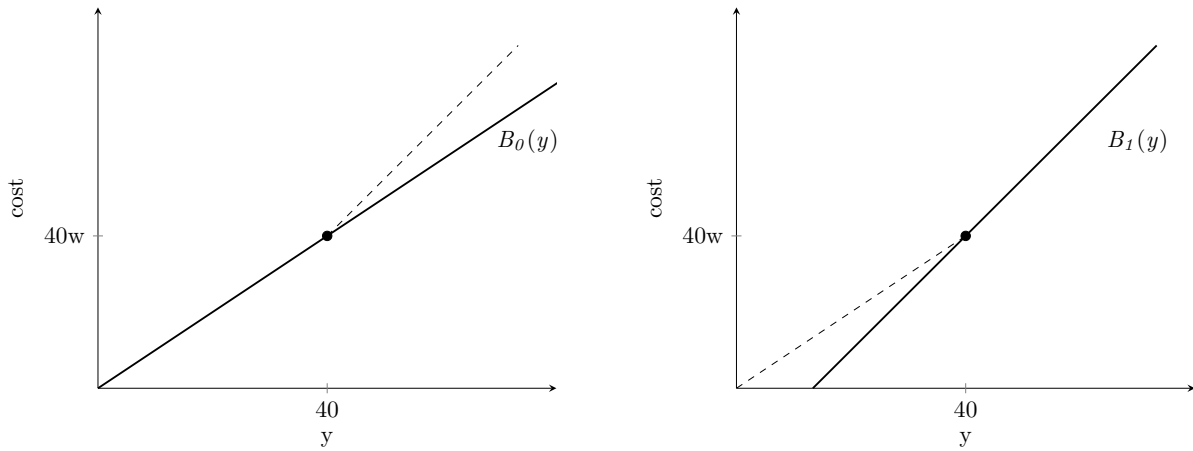
<sup>39</sup>In particular, the subgradient of  $\max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\}$  will depend on whether one approaches from the  $y_i(\mathbf{x}) > k$  or the  $y_i(\mathbf{x}) < k$  side. For example with a scalar  $x$  and linear  $B_0$  and  $B_1$ , the derivative of  $B_{ki}(x)$  discontinuously rises when  $y_i(x) = k$ .

<sup>40</sup>An example from the literature in which a distinction between  $y$  and  $\mathbf{x}$  cannot be avoided is Best et al. (2015). These authors study firms in Pakistan, who pay either a tax on output or a tax on profit, whichever is higher. The two tax schedules cross when the ratio of profits to output crosses a certain threshold that is pinned down by the two respective tax rates. In this case, the variable  $y$  depends both on production and on reported costs, leading to two margins of response to the kink: one from choosing the scale of production and the other from choosing whether and how much to misreport costs.

cost functions are:

$$B_{0i}(y) := w_i y \quad \text{and} \quad B_{1i}(y) := 1.5w_i y - 20w_i \quad (\text{A.2})$$

The functions  $B_0$  and  $B_1$  are depicted in Figure A.1 for a single worker with wage  $w_i = w$ .  $B_0$  describes a setting in which the worker is paid at their straight-time wage  $w$  for all hours, regardless of whether they work more or less than 40.  $B_1$  describes a setting in which the worker is instead paid at their overtime rate  $1.5w$  for all hours, but the firm is given a weekly “subsidy” that keeps them indifferent between the two cost schedules at  $y = 40$ . With these definitions, we can see that the actual labor cost to the firm of any number of hours  $h$  is  $B_{ki}(y) := \max\{B_{0i}(y), B_{1i}(y)\}$  for worker  $i$ .



**FIGURE A.1:** Definition of counterfactual cost functions  $B_0$  and  $B_1$  that firms could have faced, absent the overtime kink. Dashed lines show the rest of actual cost function in comparison to the counterfactual as a solid line. Note that we use the notation  $y$  here to indicate hours, rather than the  $h$  used in the main text.

## A.2 Potential outcomes as counterfactual choices

To introduce a notion of treatment effects in the bunching design, I define a pair of potential outcomes as what would occur if the decision-maker faced either of the functions  $B_0$  or  $B_1$  globally, without the kink.

**Definition (potential outcomes).** Let  $Y_{0i}$  be the value of  $y_i(\mathbf{x})$  that would occur for unit  $i$  if  $d(i)$  faced the constraint  $z \geq B_0(\mathbf{x})$ , and let  $Y_{1i}$  be the value that would occur under the constraint  $z \geq B_1(\mathbf{x})$ .

To relate these potential outcomes to choices of the decision-maker, we make explicit the assumption that they control the value of  $y_i(\mathbf{x})$ . For any function  $B$  let  $Y_{Bi}$  be the outcome that

would occur under the choice constraint  $z \geq B(\mathbf{x})$ , with  $Y_{0i}$  and  $Y_{1i}$  shorthands for  $Y_{B_{0i}i}$  and  $Y_{B_{1i}i}$ , respectively.<sup>41</sup>

**Assumption CHOICE (perfect manipulation of  $y$ ).** *For any function  $B(\mathbf{x})$ ,  $Y_{Bi} = y_i(\mathbf{x}_{Bi})$ , where  $(z_{Bi}, \mathbf{x}_{Bi})$  is the choice that  $d(i)$  would make under the constraint  $z \geq B(\mathbf{x})$ .*

Assumption CHOICE rules out for example optimization error, which could limit the decision-maker's ability to exactly manipulate values of  $\mathbf{x}$  and hence  $y$ . It also takes for granted that counterfactual choices are unique, and rules out some kinds of extensive margin effects in which a decision-maker would not choose any value of  $Y$  at all under  $B_1$  or  $B_0$ . Assumption CHOICE may be relaxed somewhat while still allowing for meaningful causal inference, but I maintain this assumption throughout (however the decision-maker need not always be the firm only; see Appendix C). Note that CHOICE here differs from the version given in the main text in that it applies to all functions  $B$ , not just  $B_0$ ,  $B_1$  and  $B_k$  (this is useful for Theorem 2).

The central behavioral assumption that allows us to reason about the counterfactuals  $Y_0$  and  $Y_1$  is that decision-makers have convex preferences over  $(c, \mathbf{x})$  and dislike costs  $z$ :

**Assumption CONVEX (strictly convex preferences, monotonic in  $z$ ).** *For each agent  $i$  and function  $B(\mathbf{x})$ , choice is  $(z_{Bi}, \mathbf{x}_{Bi}) = \operatorname{argmax}_{z, \mathbf{x}} \{u_i(z, \mathbf{x}) : z \geq B(\mathbf{x})\}$  where  $u_i(z, \mathbf{x})$  is continuous and strictly quasi-concave in  $(z, \mathbf{x})$ , and strictly decreasing in  $z$ .*

*Note:* The function  $u_i(\cdot)$  represents preferences over choice variables for unit  $i$ , but the preferences are those of the decision maker  $d(i)$ . I avoid more explicit notation like  $u_{d(i),i}(\cdot)$  for brevity. In the overtime setting with firms choosing hours,  $u_i(z, \mathbf{x})$  corresponds to the firm's profit function  $\pi$  as a function of the hours of a particular worker this week, and costs this week  $z$  for that worker.

The notation of Assumption CONVEX does not make explicit any dependence of the functions  $u_i(\cdot)$  on the choices made for other observational units  $i' \neq i$ . When the functions  $u_i(\cdot)$  are indeed invariant over such counterfactual choices, we have a version of the no-interference condition of the stable unit treatment values assumption (SUTVA). Maintaining SUTVA is not necessary to define treatment effects in the bunching design, provided that the variables  $y$  and  $z$  can be coherently defined at the individual unit  $i$  level (see Example 3 in Section 4.1). Nevertheless, the interpretation of the treatment effects identified by the bunching design is most straightforward when SUTVA does hold. This assumption is standard in the bunching design, but it may be a restrictive one in overtime context where a single firm chooses the hours of multiple workers.<sup>42</sup> I discuss this further in the overtime setting in Section 4.4 and Appendix D.

<sup>41</sup>Note that in this notation Assumption CHOICE implies that the actual outcome  $Y_i$  observed by the econometrician is equal to  $Y_{B_{ki}i}$ .

<sup>42</sup>However I note that SUTVA issues could also occur in canonical bunching designs: for example if spouses choose their labor supply jointly, the introduction of a tax kink may cause one spouse to increase labor supply while the other decreases theirs.

A weaker assumption than convexity that will still have identifying power is simply that decision-makers' choices do not violate the weak axiom of revealed preference:

**Assumption WARP (rationalizable choices).** *Consider two budget functions  $B$  and  $B'$  and any unit  $i$ . If  $d(i)$ 's choice under  $B'$  is feasible under  $B$ , i.e.  $z_{B'i} \geq B(\mathbf{x}_{B'i})$ , then  $(z_{Bi}, \mathbf{x}_{Bi}) = (z_{B'i}, \mathbf{x}_{B'i})$ .*

I make the stronger assumption CONVEX for most of the identification results, but Assumption WARP still allows a version of many of them in which equalities become weak inequalities, indicating a degree of robustness with respect to departures from convexity. Note that the monotonicity assumption in CONVEX implies that choices will always satisfy  $z = B(\mathbf{x})$ , i.e. agents' choices will lay on their cost functions (despite Eq. A.1 being an inequality, indicating "free-disposal").

In the overtime application, the potential outcomes  $Y_{0i}$  and  $Y_{1i}$  are the hours that the firm would choose, respectively, in a situation a) in which there was no overtime premium and the firm always had to pay  $w_i$  for each hour; and b) a situation in which the firm were to pay  $1.5w_i$  for all hours of labor, but receive a subsidy of  $20w_i$  that keeps the firm indifferent between  $B_0$  and  $B_1$  when  $h = 40$  (cf. Eq. A.2). When firm preferences are quasilinear with respect to wage costs, the choice of hours  $Y_1$  will be the same as what the firm would have chosen without the subsidy of  $20w$ .

### A.3 Examples from the general choice model in the overtime setting

To demonstrate the added flexibility of the general choice model, I below present some examples in the overtime context. These examples are illustrative, and each could apply to a different subset of units in the population. In these examples we continue to take the decision-maker for a given unit  $it$  to be the firm employing worker  $i$ .<sup>43</sup>

#### *Example 1: Substitution from bonus pay*

Let the firm's choice vector be  $\mathbf{x} = (h, b)'$ , where  $b \geq 0$  indicates a bonus (or other fringe benefit) paid to the worker. Firms may find it optimal to offer bonuses to improve worker satisfaction and reduce turnover. Suppose firm preferences are:  $\pi(z, h, b) = f(h) + g(z + b - \nu(h)) - z - b$ , where  $z$  continues to denote wage compensation this week,  $z + b - \nu(h)$  is the worker's utility with  $\nu(h)$  a convex disutility from labor  $h$ , and  $g(\cdot)$  increasing and concave. In this model firms will choose the surplus maximizing choice of hours  $h_m := \arg\max_h f(h) - \nu(h)$ , provided that the corresponding optimal bonus is non-negative. Bonuses fully adjust to counteract overtime costs, and  $h_0 = h_1 = h_m$ .

---

<sup>43</sup>Appendix C discusses a further example in which the firm and worker bargain over this week's hours. This model can attenuate the wage elasticity of chosen hours since overtime pay gives the parties opposing incentives.

*Example 2: Off-the-clock hours and paid breaks*

Suppose firms choose a pair  $\mathbf{x} = (h, o)'$  with  $h$  hours worked and  $o$  hours worked “off-the-clock”, such that  $y(\mathbf{x}) = h - o$  are the hours for which the worker is ultimately paid. Evasion is harder the larger  $o$  is, which could be represented by firms facing a convex evasion cost  $\phi(o)$ , so that firm utility is  $\pi(z, h, o) = f(h) - \phi(o) - z$ .<sup>44</sup> This model can also include some firms voluntarily offering paid breaks by allowing  $o$  to be negative.

*Example 3: Complementaries between workers or weeks*

Suppose the firm simultaneously chooses the hours  $\mathbf{x} = (h, g)$  of two workers according to production that is iso-elastic in a CES aggregate ( $g$  could also denote planned hours next week):  $\pi(z, h, g) = a \cdot ((\gamma h^\rho + g^\rho)^{1/\rho})^{1+\frac{1}{\epsilon}} - z$  with  $\gamma$  a relative productivity shock. Let  $g^*$  denote the firm’s optimal choice of hours for the second worker. Optimal  $h$  then maximizes  $\pi(z, h, g^*)$  subject to  $z = B_k(h)$ , as if the firm faced a single-worker production function of  $f(h) = a \cdot ((\gamma h^\rho + g^{*\rho})^{1/\rho})^{1+\frac{1}{\epsilon}}$ . This function is more elastic than  $a \cdot h^{1+\frac{1}{\epsilon}}$  provided that  $\rho < 1 + 1/\epsilon$ , attenuating the response to an increase in  $w$  implied by a given  $\epsilon$ .<sup>45</sup> Section 4.4 discusses how complementaries affect the final evaluation of the FLSA.

**Further notes on the general model**

I conclude this section with some further remarks on the generality of Eq. (A.1) given the above assumptions. The first is that the budget functions  $B_0$  and  $B_1$  can depend on a subset of the variables that enter into the function for  $y$ , and vice versa. In the former case, this is because the only restriction on  $B_{0i}(\mathbf{x})$  and  $B_{1i}(\mathbf{x})$  is that they are continuous and weakly convex in all components of  $\mathbf{x}$ ; thus, having zero dependence on a component of  $\mathbf{x}$  is permissible. This is of particular interest because while the variables entering into the budget functions are generally known from the empirical context generating the kink, the model can allow additional choice variables to enter into the threshold-crossing variable  $y$ , that may not even be known to the econometrician. Beyond the examples above for the overtime setting,

Supplemental Material Section 2 details an example from the literature (Best et al., 2015) in which the functions  $B_0$  and  $B_1$  depend directly on two components of  $\mathbf{x}$ . In general, we might suppose that  $B_{di}(\mathbf{x}) = B_{di}(\bar{\mathbf{x}})$ , where  $\bar{\mathbf{x}}$  is a sub-vector of the first  $n$  components of  $\mathbf{x}$ , but  $y_i(\mathbf{x})$

---

<sup>44</sup>Note that the data observed in our sample are of hours of work  $y(\mathbf{x})$  for which the worker is paid, when this differs from  $h$ . Appendix A describes how Equation 2 still holds, but for counterfactual values of hours *paid*  $y = h - o$  rather than hours worked  $h$ . The bunching design lets us investigate treatment effects on paid hours, without observing off-the-clock hours or break time  $o$ .

<sup>45</sup>This expression overstates the degree of attenuation somewhat, since  $h_1$  and  $h_0$  maximize  $f(h)$  above for different values  $g^*$ , which leads to a larger gap between  $h_0$  and  $h_1$  compared with a fixed  $g^*$  by the Le Chatelier principle (Milgrom and Roberts, 1996). However  $h_1/h_0$  still increases on net given  $\rho < 1 + 1/\epsilon$ .

is still a function of all  $n + l$  components of  $\mathbf{x}$ . The values of the remaining  $l$  components affect the decision-maker's optimizing choice of  $y$ , because they affect the value of  $y$  and hence which regime  $B_0$  or  $B_1$  the decision-maker's choice is in. Thus, observed bunching in  $y$  can reflect a response along any of these  $l$  additional margins, even though they correspond to variables that are unobserved or even unknown to the researcher. This can complicate identification of specific structural elasticities, but does not challenge the credibility of causal inference about  $y$  alone.

#### A.4 Observables in the kink bunching design

Lemma 1 outlines the core consequence of convexity of preferences for the relationship between observed  $Y_i$  and the potential outcomes introduced in the last section:

**Lemma 1 (realized choices as truncated potential outcomes).** *Under Assumptions CONVEX and CHOICE, the outcome observed given the constraint  $z \geq \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\}$  is:*

$$Y_i = \begin{cases} Y_{0i} & \text{if } Y_{0i} < k \\ k & \text{if } Y_{1i} \leq k \leq Y_{0i} \\ Y_{1i} & \text{if } Y_{1i} > k \end{cases}$$

*Proof.* See Appendix B. □

Lemma 1 says that the pair of counterfactual outcomes  $(Y_{0i}, Y_{1i})$  is sufficient to pin down actual choice  $Y_i$ , which can in fact be seen as an observation of one or the other potential outcome depending on how they relate to the kink point  $k$ . When the  $Y_{0i}$  potential outcome is greater than  $k$  but the  $Y_{1i}$  potential outcome is below – when the potential outcomes “straddle” the kink – the decision will locate choose the corner solution of locating exactly the kink.<sup>46</sup>

Note that the straddling event from Lemma 1 can be written as  $Y_{0i} \in [k, k + \Delta_i]$ , where  $\Delta_i = Y_{0i} - Y_{1i}$ . Equivalently, we could also write  $Y_{1i} \leq k \leq Y_{0i}$  as  $Y_i \in [k - \Delta_i, k]$ . This forms the basic link between bunching and treatment effects, where  $\Delta_i$  can be thought of as the causal effect of a counterfactual change from the choice set under  $B_1$  to the choice set under  $B_0$ . Let  $\mathcal{B} := P(Y_i = k)$  be the observable probability that the decision-maker chooses to locate exactly at  $Y = k$ . This relationship holds in a weakened form under WARP rather than CONVEX:

**Proposition 1 (relation between bunching and straddling).** *a) Under CONVEX and CHOICE:  $\mathcal{B} = P(Y_{0i} \in [k, k + \Delta_i])$ ; b) under WARP and CHOICE:  $\mathcal{B} \leq P(Y_{0i} \in [k, k + \Delta_i])$ .*

*Proof.* See Appendix B. □

---

<sup>46</sup>The opposite situation of  $Y_{0i} \leq k \leq Y_{1i}$ , what we might call “reverse straddling”, is ruled out by WARP when it occurs by at least one strict inequality.

Consider a random sample of observations of  $Y_i$ . Under i.i.d. sampling of  $Y_i$ , the distribution  $F(y)$  of  $Y_i$  is identified.<sup>47</sup> Let  $F_0(y) = P(Y_{0i} \leq y)$  be the distribution function of the random variable  $Y_0$ , and  $F_1(y)$  the distribution function of  $Y_1$ . From Lemma 1 it follows immediately that  $F_0(y) = F(y)$  for all  $y < k$ , and  $F_1(y) = F(y)$  for  $y > k$ . Thus observations of  $Y_i$  are also informative about the marginal distributions of  $Y_{0i}$  and  $Y_{1i}$ . Again, a weaker version of this also holds under WARP rather than CONVEX:

**Proposition 2 (identification of truncated densities).** *Suppose that  $F_0$  and  $F_1$  are continuously differentiable with derivatives  $f_0$  and  $f_1$ , and that  $F$  admits a derivative function  $f(y)$  for  $y \neq k$ . Under WARP and CHOICE:  $f_0(y) \leq f(y)$  for  $y < k$  and  $f_0(k) \leq \lim_{y \uparrow k} f(y)$ , while  $f_1(y) \leq f(y)$  for  $y > k$  and  $f_1(k) \leq \lim_{y \downarrow k} f(y)$ , with equalities under CONVEX.*

*Proof.* See Appendix B. □

## A.5 Treatment effects in the bunching design

Proposition 1 establishes that bunching can be informative about features of the distribution of treatment effects  $\Delta_i$ . This section discusses the interpretation of these treatment effects and some additional identification results omitted in the main text.

### *Discussion of treatment effects vs. structural parameters:*

The treatment effects  $\Delta_i$  are “reduced form” in the sense that when the decision-maker has multiple margins of response  $\mathbf{x}$  to the incentives introduced by the kink, these may be bundled together in the treatment effect  $\Delta_i$ . This clarifies a limitation sometimes levied against the bunching design, while also revealing a perhaps under-appreciated strength. On the one hand, it is not always clear “which elasticity” is elicited by bunching at a kink, complicating efforts to identify a elasticity parameter having a firm structural interpretation.

On the other hand, the bunching design can be useful for ex-post policy evaluation and even forecasting effects of small policy changes (as described in Section 4.4), without committing to a tightly parameterized underlying model of choice. This provides a response to Einav et al. (2017), who caution that alternative structural models calibrated from the bunching-design can yield very different predictions about counterfactuals. By focusing on the counterfactuals  $Y_{0i}$  and  $Y_{1i}$ , we can specify a *particular* type of counterfactual question that can be answered robustly across a broad class of models.

The “trick” of Lemma 1 is to express the observable data in terms of counterfactual choices, rather than of primitives of the utility function. The underlying utility function  $u_i(z, \mathbf{x})$  is used

---

<sup>47</sup>Note that in the overtime application sampling is actually at the firm level, which coincides with the level of decision-making units  $d(i)$ .

only as an intermediate step, only requiring the nonparametric restrictions of convexity and monotonicity rather than knowing its functional form. The econometrician need not even know the full vector  $\mathbf{x}$  of choice variables underlying agents' observed value of  $y$ , they simply need to believe that preferences are convex in them, and verify that  $B_0$  and  $B_1$  are convex in a subset of them. This greatly increases the robustness of the method to potential misspecification of the underlying choice model. The Supplemental Material illustrates this point looking at the setting of Best et al. (2015).

*Additional identification results for the bunching design:*

While Theorem 1 of Section 4 develops the treatment effect identification result used to evaluate the FLSA, Supplemental Appendix 2 presents some further identification results for the bunching design that are not used in this paper, which can be considered alternatives to Theorem 1. This includes re-expressing various results in the general framework of this section, including the linear interpolation approach of Saez (2010), the polynomial approach of Chetty et al. (2011) and a “small-kink” approximation appearing in Saez (2010) and Kleven (2016). The Supplemental Appendix also outlines alternative shape constraints to bi-log-concavity, including monotonicity of densities. I also give there a result in which a lower bound to a certain local average treatment effect is identified under WARP, without requiring convexity of preferences.

*The buncher ATE when Assumption RANK fails:*

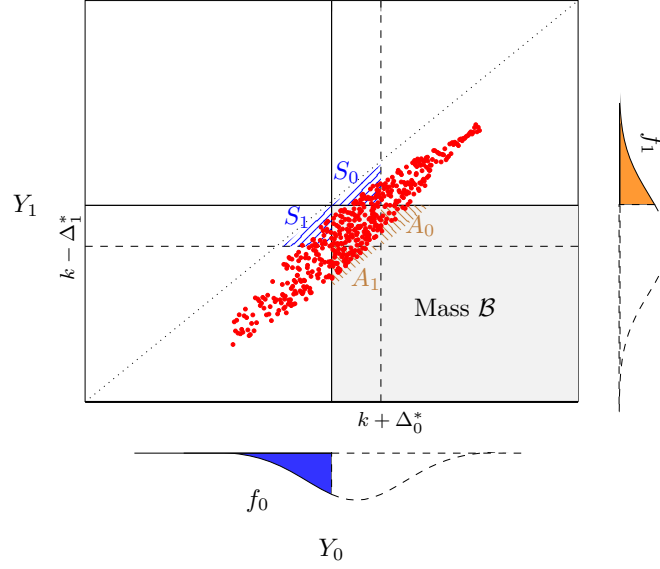
This section picks up from the discussion in Section 4.3, which introduces the buncher ATE  $\Delta_k^*$  parameter and Assumption RANK, but continues with the notation of this Appendix. When RANK fails (and  $p = 0$  for simplicity), the bounds from Theorem 1 are still valid for the averaged quantile treatment effect:

$$\frac{1}{B} \int_{F_0(k)}^{F_1(k)} Q_0(u) - Q_1(u) = \mathbb{E}[Y_{0i} | Y_{0i} \in [k, k + \Delta_0^*]] - \mathbb{E}[Y_{1i} | Y_{1i} \in [k - \Delta_1^*, k]] \quad (\text{A.3})$$

under BLC of  $Y_0$  and  $Y_1$ , where we define  $\Delta_0^* := Q_0(F_1(k)) - Q_1(F_1(k)) = Q_0(F_1(k)) - k$  and  $\Delta_1^* := Q_0(F_0(k)) - Q_1(F_0(k)) = k - Q_1(F_0(k))$ . This can be seen to yield a lower bound on the buncher ATE, as described in Figure A.2 below.



### Signing the bias when RANK fails



**FIGURE A.2:** When Assumption RANK fails, the average  $E[Y_{0i}|Y_{0i} \in [k, k + \Delta_0^*]]$  will include the mass in the region  $S_0$ , who are not bunchers (NE lines) but will be missing the mass in the region  $A_0$  (NW lines) who are. This causes an under-estimate of the desired quantity  $E[Y_{0i}|Y_{1i} \leq k \leq Y_{0i}]$ . Similarly,  $E[Y_{1i}|Y_{1i} \in [k - \Delta_1^*, k]]$  will include the mass in the region  $S_1$ , who are not bunchers but will be missing the mass in  $A_1$ , who are. This causes an over-estimate of the desired quantity  $E[Y_{1i}|Y_{1i} \leq k \leq Y_{0i}]$ .

## A.6 Policy changes in the bunching-design

Consider a bunching design in which the cost functions  $B_0$  and  $B_1$  can be viewed as members of family  $B_i(\mathbf{x}; \rho, k)$  parameterized by a continuum of scalars  $\rho$  and  $k$ , where  $B_{0i}(\mathbf{x}) = B_i(\mathbf{x}; \rho_0, k^*)$  and  $B_{1i}(\mathbf{x}) = B_i(\mathbf{x}; \rho_1, k^*)$  for some  $\rho_1 > \rho_0$  and value  $k^*$  of  $k$ . In the overtime setting  $\rho$  represents a wage-scaling factor, with  $\rho = 1$  for straight-time and  $\rho = 1.5$  for overtime:

$$B_i(y; \rho, k) = \rho w_i y - k w_i (\rho - 1) \quad (\text{A.4})$$

where work hours  $y$  may continue to be a function  $y(\mathbf{x})$  of a vector of choice variables to the firm. Here  $\rho$  represents an arbitrary wage-scaling factor, while  $k$  controls the size of a lump-sum subsidy that keeps  $B_i(k; \rho, k)$  invariant across  $\rho$ .

Assume that  $\rho$  takes values in a convex subset of  $\mathbb{R}$  containing  $\rho_0$  and  $\rho_1$ , and that for any  $k$  and  $\rho' > \rho$  the cost functions  $B_i(\mathbf{x}; \rho, k)$  and  $B_i(\mathbf{x}; \rho', k)$  satisfy the conditions of the bunching design framework from Section 4, with the function  $y_i(\mathbf{x})$  fixed across all such values. That is,  $B_i(\mathbf{x}; \rho', k) > B_i(\mathbf{x}; \rho, k)$  iff  $y_i(\mathbf{x}) > k$  with equality when  $y_i(\mathbf{x}) = k$ , the functions  $B_i(\cdot; \rho, k)$  are weakly convex and continuous, and  $y_i(\cdot)$  is continuous. It is readily verified that Equation (A.4)

satisfies these requirements with  $y_i(h) = h$ .<sup>48</sup>

For any value of  $\rho$ , let  $Y_i(\rho, k)$  be agent  $i$ 's realized value of  $y_i(\mathbf{x})$  when a choice of  $(z, \mathbf{x})$  is made under the constraint  $c \geq B_i(\mathbf{x}; \rho, k)$ . A natural restriction in the overtime setting that is that the function  $Y_i(\rho, k)$  does not depend on  $k$ , and some of the results below will require this. A sufficient condition for  $Y_i(\rho, k) = Y_i(\rho)$  is a family of cost functions that are linearly separable in  $k$ , as we have in Equation (A.4), along with quasi-linearity of preferences:

**Assumption SEPARABLE (invariance of potential outcomes with respect to  $k$ ).** *For all  $i, \rho$  and  $k$ ,  $B_i(\mathbf{x}; \rho, k)$  is additively separable between  $k$  and  $\mathbf{x}$  (e.g.  $b_i(\mathbf{x}, \rho) + \phi_i(\rho, k)$  for some functions  $b_i$  and  $\phi_i$ ), and for all  $i$   $u_i(z, \mathbf{x})$  can be chosen to be additively separable and linear in  $z$ .*

Quasilinearity of preferences is a property of profit-maximizing firms when  $c$  represents a cost, thus it is a natural assumption in the overtime setting. However, additive separability of  $B(\mathbf{x}; \rho, k)$  in  $k$  may be context specific: in the example from Best et al. (2015) described in Appendix A, quasi-linearity of preferences is not sufficient since the cost functions are not additively separable in  $k$ . To maintain clarity of exposition, I will keep  $k$  implicit in  $Y_i(\rho)$  throughout the foregoing discussion, but the proofs make it clear when SEPARABLE is being used.

Below I state two intermediate results that allow us to derive expressions for the effects of marginal changes to  $\rho_1$  or  $k$  on hours. Lemma 2 generalizes an existing result from Blomquist et al. (2021), and makes use of a regularity condition I introduce in the proof as Assumption SMOOTH.<sup>49</sup> Counterfactual bunchers  $K_i^* = 1$  are assumed to stay at  $k^*$ , regardless of  $\rho$  and  $k$ . Let  $p(k) = p \cdot \mathbb{1}(k = k^*)$  denote the possible counterfactual mass at the kink as a function of  $k$ . Let  $f_\rho(y)$  be the density of  $Y_i(\rho)$ , which exists by SMOOTH and is defined for  $y = k^*$  as a limit (see proof).

**Lemma 2 (bunching from marginal responsiveness).** *Assume CHOICE, SMOOTH and WARP. Then:*

$$\mathcal{B} - p(k) \leq \int_{\rho_0}^{\rho_1} f_\rho(k) \mathbb{E} \left[ -\frac{dY_i(\rho)}{d\rho} \middle| Y_i(\rho) = k \right] d\rho$$

*with equality under CONVEX.*

*Proof.* See Supplemental Material. □

The main tool in establishing Lemma 2 is relate the integrand in the above to the rate at which kink-induced bunching goes away as the “size” of the kink goes to zero.

<sup>48</sup>As an alternative example, I construct in Appendix A functions  $B_i(\mathbf{x}; \rho, k)$  for the bunching design setting from Best et al. (2015). In that case,  $\rho$  parameterizes a smooth transition between an output and a profit tax, where  $k$  enters into the rate applied to the tax base for that value of  $\rho$ .

<sup>49</sup>Blomquist et al. (2021) derive the special case of Lemma 2 with CONVEX and  $p = 0$ , in the context of a more restricted model of labor supply under taxation. I establish it here for the general bunching design model where in particular, the  $Y_i(\rho)$  may depend on an underlying vector  $\mathbf{x}$  which are not observed by the econometrician. I also use different regularity conditions.

**Lemma SMALL (small kink limit).** Assume *CHOICE\**, *WARP*, and *SMOOTH*. Then:

$$\lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho') \leq k \leq Y_i(\rho)) - p(k)}{\rho' - \rho} = -f_\rho(k) \mathbb{E} \left[ \frac{dY_i(\rho)}{d\rho} \middle| Y_i(\rho) = k \right]$$

*Proof.* See Supplemental Material. □

Note that the quantity  $P(Y_i(\rho') \leq k \leq Y_i(\rho)) - p(k)$  is an upper bound on the bunching that would occur due to a kink between budget functions  $B_i(\mathbf{x}; \rho, k)$  and  $B_i(\mathbf{x}; \rho', k)$  (with equality under CONVEX). Lemma SMALL thus shows that the small-kink approximation that has appeared in Saez (2010) and Kleven (2016) (stated in Theorem 6 of the Supplemental Material) becomes exact in the limit that the two cost functions approach one another, since treatment effects are:

$$Y_i(\rho) - Y_i(\rho') = \frac{dY_i(\rho)}{d\rho}(\rho' - \rho) + O((\rho' - \rho)^2)$$

By Lemma 2, we can also see that the RHS in Lemma SMALL evaluated at  $\rho = \rho_1$  is equal to the derivative of bunching as  $\rho_1$  is increased, under CONVEX.

Lemma 2 is particularly useful when combined with a result from Kasy (2017), which considers how the distribution of a generic outcome variable changes as heterogeneous units flow to different values of that variable in response to marginal policy changes.

**Lemma 3 (flows under a small change to  $\rho$ ).** Under *SMOOTH*:

$$\partial_\rho f_\rho(y) = \partial_y \left\{ f_\rho(y) \mathbb{E} \left[ -\frac{dY_i(\rho)}{d\rho} \middle| Y_i(\rho) = y, K_i^* = 0 \right] \right\}$$

*Proof.* See Supplemental Material or Kasy (2017) for original result. □

The intuition behind Lemma 3 comes from the physical dynamics of fluids. When  $\rho$  changes, a mass of units will “flow” out of a small neighborhood around any  $y$ , and this mass is proportional to the density at  $y$  and to the average rate at which units move in response to the change. When the magnitude of this net flow varies with  $y$ , the change to  $\rho$  will lead to a change in the density there.

With  $\rho_0$  fixed at some value, let us index observed  $Y_i$  and bunching  $\mathcal{B}$  with the superscript  $[k, \rho_1]$  when they occur in a kinked policy environment with cost functions  $B_i(\cdot; \rho_0, k)$  and  $B_i(\cdot; \rho_1, k)$ . Lemmas 2 and 3 together imply Theorem 2, which in the notation of this section reads as:

1.  $\partial_k \left\{ \mathcal{B}^{[k, \rho_1]} - p(k) \right\} = f_1(k) - f_0(k)$
2.  $\partial_k \mathbb{E}[Y_i^{[k, \rho_1]}] = \mathcal{B}^{[k, \rho_1]} - p(k)$
3.  $\partial_{\rho_1} \mathcal{B}^{[k, \rho_1]} = -k f_{\rho_1}(k) \mathbb{E} \left[ \frac{dY_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = k \right]$

$$4. \partial_{\rho_1} \mathbb{E}[Y_i^{[k, \rho_1]}] = - \int_k^\infty f_{\rho_1}(y) \mathbb{E} \left[ \frac{dY_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = y \right] dy$$

*Note:* Assumption SEPARABLE is only necessary for Items 1-2 in Theorem 2, Item 3 holds without it and with  $\frac{\partial Y_i(\rho, k)}{\partial \rho}$  replacing  $\frac{dY_i(\rho)}{d\rho}$ .

## B Main Proofs

### B.1 Proof of Lemma 1

The proof proceeds in the following two steps:

- i) First, I show that  $Y_{0i} \leq k$  implies that  $Y_i = Y_{0i}$ , and similarly  $Y_{1i} \geq k$  implies that  $Y_i = Y_{1i}$ . This holds under CONVEX but also under the weaker assumption of WARP.
- ii) Second, I show that under CONVEX  $Y_i < k \implies Y_i = Y_{0i}$  and  $Y_i > k \implies Y_i = Y_{1i}$ .

Item i) above establishes the first and third cases of Lemma 1. The only remaining possible case is that  $Y_{1i} \leq k \leq Y_{0i}$ . However, to finish establishing Lemma 1, we also need the reverse implication: that  $Y_{1i} \leq k \leq Y_{0i}$  implies  $Y_i = k$ . This comes from taking the contrapositive of each of the two claims in item ii).

**Proof of i):** Let  $\mathcal{X}_{0i} = \{\mathbf{x} : y_i(\mathbf{x}) \leq k\}$  and  $\mathcal{X}_{1i} = \{\mathbf{x} : y_i(\mathbf{x}) \geq k\}$ . If  $Y_{0i} \leq k$ , then by CHOICE  $\mathbf{x}_{B_0}$  is in  $\mathcal{X}_{0i}$ . Since  $B_k(\mathbf{x}) = B_0(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{X}_{0i}$ , it follows that  $z_{B_0i} \geq B_k(\mathbf{x}_{B_0i})$ , i.e.  $Y_{0i}$  is feasible under  $B_k$ . Note that  $B_{ki}(\mathbf{x}) \geq B_{0i}(\mathbf{x})$  for all  $\mathbf{x}$ . By WARP then  $(z_{B_{ki}}, \mathbf{x}_{B_{ki}}) = (z_{B_{0i}}, \mathbf{x}_{B_{0i}})$ . Thus  $Y_i = y_i(\mathbf{x}_{B_k}) = y_i(\mathbf{x}_{B_0}) = Y_{0i}$ . So  $Y_{0i} \leq k \implies Y_i = Y_{0i}$ . By the same logic we can show that  $Y_{1i} \geq k \implies Y_i = Y_{1i}$ .

**Proof of ii):** For any convex budget function  $B(\mathbf{x})$ ,  $(z_{Bi}, \mathbf{x}_{Bi}) = \operatorname{argmax}_{z, \mathbf{x}} \{u_i(z, \mathbf{x}) \text{ s.t. } z \geq B(\mathbf{x})\}$  where the RHS exists and is unique since it maximizes the strictly quasi-concave function  $u_i(z, \mathbf{x})$  over the convex domain  $\{(z, \mathbf{x}) : z \geq B(\mathbf{x})\}$ . Furthermore, by monotonicity of  $u(z, \mathbf{x})$  in  $z$  we may substitute in the constraint  $z = B(\mathbf{x})$  and write

$$\mathbf{x}_{Bi} = \operatorname{argmax}_{\mathbf{x}} u_i(B(\mathbf{x}), \mathbf{x})$$

Consider any  $\mathbf{x} \neq \mathbf{x}_{Bi}$ , and let  $\tilde{\mathbf{x}} = \theta \mathbf{x} + (1 - \theta) \mathbf{x}^*$  where  $\mathbf{x}^* = \mathbf{x}_{Bi}$  and  $\theta \in (0, 1)$ . Since  $B(\mathbf{x})$  is convex in  $\mathbf{x}$  and  $u_i(z, \mathbf{x})$  is weakly decreasing in  $z$ :

$$u_i(B(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}) \geq u_i(\theta B(\mathbf{x}) + (1 - \theta) B(\mathbf{x}^*), \tilde{\mathbf{x}}) > \min\{u_i(B(\mathbf{x}), \mathbf{x}), u_i(B(\mathbf{x}^*), \mathbf{x}^*)\} = u_i(B(\mathbf{x}), \mathbf{x}) \quad (\text{B.5})$$

where I have used strict quasi-concavity of  $u_i(z, \mathbf{x})$  in the second step, and that  $\mathbf{x}^*$  is a maximizer in the third. This result implies that for any  $\mathbf{x} \neq \mathbf{x}^*$ , if one draws a line between  $\mathbf{x}$  and  $\mathbf{x}^*$ , the function  $u_i(B(\mathbf{x}), \mathbf{x})$  is strictly increasing as one moves towards  $\mathbf{x}^*$ . When  $\mathbf{x}$  is a scalar, this argument is used by Blomquist et al. (2015) (see Lemma A2 therein) to show that  $u_i(B(\mathbf{x}), \mathbf{x})$  is strictly increasing to the left of  $\mathbf{x}^*$ , and strictly decreasing to the right of  $\mathbf{x}^*$ . Note that for any (binding) linear budget constraint  $B(\mathbf{x})$ , the result holds without monotonicity of  $u_i(z, \mathbf{x})$  in  $z$ . This is useful for Theorem 2\* in which some workers choose their hours.

For any function  $B$ , let  $u_{B_i}(\mathbf{x}) = u_i(B(\mathbf{x}), \mathbf{x})$ , and note that

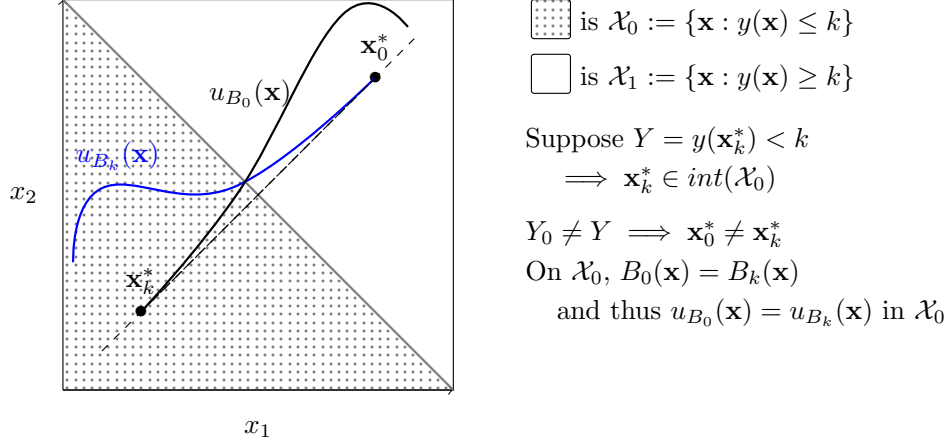
$$u_{B_k i}(\mathbf{x}) = \begin{cases} u_{B_{0i}}(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{X}_{0i} \\ u_{B_{1i}}(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{X}_{1i} \end{cases}$$

Let  $\mathbf{x}_{ki}$  be the unique maximizer of  $u_{B_k i}(\mathbf{x})$ , where  $Y_i = y_i(\mathbf{x}_{ki})$ . Suppose that  $Y_i < k$ . By continuity of  $y_i(\mathbf{x})$ ,  $\mathcal{X}_{0i}$  is a closed set and  $\mathbf{x}_{ki}$  belongs to the interior of  $\mathcal{X}_{0i}$ . Suppose furthermore that  $Y_{0i} \neq Y_i$ , with  $\mathbf{x}_{0i}$  the maximizer of  $u_{B_{0i}}(\mathbf{x})$ . If this were the case, then there would exist a point  $\tilde{\mathbf{x}} \in \mathcal{X}_{0i}$  along the line from  $\mathbf{x}_{0i}$  to  $\mathbf{x}_{ki}$ . By Eq. (B.5) with  $B = B_k$ , we must have  $u_{B_k i}(\tilde{\mathbf{x}}) > u_{B_k i}(\mathbf{x}_{0i})$ . Since  $u_{B_{0i}}(\mathbf{x}) = u_{B_k i}(\mathbf{x})$  in  $\mathcal{X}_{0i}$  this means that  $u_{B_{0i}}(\tilde{\mathbf{x}}) > u_{B_{0i}}(\mathbf{x}_{0i})$ , contradicting the premise that  $\mathbf{x}_{0i}$  maximizes  $u_{B_{0i}}(\mathbf{x})$ . Figure B.3 depicts the logic visually. Thus,  $Y_i < k$  implies  $Y_i = Y_{0i}$ . The proof that  $Y_i > k$  implies  $Y_i = Y_{1i}$  is analogous.

## B.2 Proof of Propositions 1 and 2

Consider Proposition 1. Item i) in the proof of Lemma 1 establishes that under CHOICE and WARP  $Y_i = k$  implies  $Y_{1i} \leq k \leq Y_{0i}$ , since taking contrapositives we have that  $(Y_i \geq k \text{ and } Y_i \leq k)$  implies  $Y_{1i} \leq k \leq Y_{0i}$ . We have also seen from item ii) that under CHOICE and CONVEX  $Y_{1i} \leq k \leq Y_{0i}$  also implies  $Y_i = k$ , thus  $Y_{1i} \leq k \leq Y_{0i}$  and  $Y_i = k$  are equivalent. Note that by adding  $\Delta_{0i} = Y_{0i} - Y_{1i}$  to both sides of the inequality  $Y_{1i} \leq k$ , we have that  $Y_{0i} \leq k + \Delta_i$ . Combining with the other inequality that  $Y_{0i} \geq k$ , we can thus rewrite the event  $Y_{1i} \leq k \leq Y_{0i}$  as  $Y_{0i} \in [k, k + \Delta_i]$  (or equivalently to  $Y_{1i} \in [k - \Delta_{0i}, k]$ ). We thus have that  $\mathcal{B} \leq P(Y_{0i} \in [k, k + \Delta])$  under CHOICE and WARP, and that  $\mathcal{B} = P(Y_i = k) = P(Y_{1i} \leq k \leq Y_{0i})$  under CHOICE and CONVEX.

Now consider Proposition 2. By item i) in the proof of Proposition 1, we have that under WARP and CHOICE  $Y_{0i} \leq k \implies Y_i = Y_{0i}$ . Thus, for any  $\delta > 0$  and  $y < k$ :  $Y_{0i} \in [y - \delta, y]$  implies that  $Y_i \in [y - \delta, y]$  and hence  $P(Y_{0i} \in [y - \delta, y]) \leq P(Y_i \in [y - \delta, y])$ . This implies that  $f_0(y) - f(y) = \lim_{\delta \downarrow 0} \frac{P(Y_{0i} \in [y - \delta, y]) - P(Y_i \in [y - \delta, y])}{\delta} \leq 0$ , i.e. that  $f(y) \geq f_0(y)$ . An analogous argument holds for  $Y_{1i}$ , where we consider the event  $Y_{1i} \in [y, y + \delta]$  any  $y > k$ . Under CONVEX



**FIGURE B.3:** Depiction of the step establishing  $(Y < k) \implies (Y = Y_0)$  in the proof of Lemma 1. In this example  $z = (x_1, x_2)$  and  $y(\mathbf{x}) = x_1 + x_2$ . We suppress indices  $i$  for clarity. Proof is by contradiction. If  $Y_0 \neq Y$ , then  $\mathbf{x}_k^* \neq \mathbf{x}_0^*$ , where  $\mathbf{x}_k^*$  and  $\mathbf{x}_0^*$  are the unique maximizers of  $u_{B_k}(\mathbf{x})$  and  $u_{B_0}(\mathbf{x})$ , respectively. By Equation B.5, we have that the function  $u_{B_0}(\mathbf{x})$ , depicted heuristically as a solid black curve, is strictly increasing as one moves along the dotted line from  $\mathbf{x}_k^*$  towards  $\mathbf{x}_0^*$ . Similarly, the function  $u_{B_k}(\mathbf{x})$ , depicted as a solid blue curve, is strictly increasing as one moves in the opposite direction along the same line, from  $\mathbf{x}_0^*$  towards  $\mathbf{x}_k^*$ . By the assumption that  $Y < k$ , then using continuity of  $y(\mathbf{x})$  it must be the case that  $\mathbf{x}_k^*$  lies in the interior of  $\mathcal{X}_0$ , the set of  $\mathbf{x}$ 's that make  $y(\mathbf{x}) \leq k$ . This means that there is some interval of the dotted line that is within  $\mathcal{X}_0$  (regardless of whether  $\mathbf{x}_0^*$  is also within  $\mathcal{X}_0$ , or it is not, as depicted). On this interval, the functions  $B_0$  and  $B_k$  are equal, and thus so must be the functions  $u_{B_k}$  and  $u_{B_0}$ . Since the same function cannot be both strictly increasing and strictly decreasing, we have obtained a contradiction.

instead of WARP, the inequalities are all equalities, by Lemma 1.

### B.3 Proof of Theorem 1

Theorem 1 of Dömbgen et al. (2017) gives a characterization of bi-log concavity in terms of a random variable's CDF and its density. In our case this reads as follows: for  $d \in \{0, 1\}$  and any  $t$ ,

$$1 - (1 - F_{d|K^*=0}(k))e^{-\frac{f_{d|K^*=0}(k)}{1 - F_{d|K^*=0}(k)}t} \leq F_{d|K^*=0}(k + t) \leq F_{d|K^*=0}(k)e^{\frac{f_{d|K^*=0}(k)}{F_{d|K^*=0}(k)}t}$$

Defining  $u = F_{0|K^*=0}(k + t)$ , we can use the substitution  $t = Q_{0|K^*=0}(u) - k$  to translate the above into bounds on the conditional quantile function of  $Y_{0i}$ , evaluated at  $u$ :

$$\frac{F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)} \cdot \ln \left( \frac{u}{F_{0|K^*=0}(k)} \right) \leq Q_{0|K^*=0}(u) - k \leq -\frac{1 - F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)} \cdot \ln \left( \frac{1 - u}{1 - F_{0|K^*=0}(k)} \right)$$

And similarly for  $Y_1$ , letting  $v = F_{1|K^*=0}(k - t)$ :

$$\frac{1 - F_{1|K^*=0}(k)}{f_{1|K^*=0}(k)} \cdot \ln \left( \frac{1 - v}{1 - F_{1|K^*=0}(k)} \right) \leq k - Q_{1|K^*=0}(v) \leq -\frac{F_{1|K^*=0}(k)}{f_{1|K^*=0}(k)} \cdot \ln \left( \frac{v}{F_{1|K^*=0}(k)} \right)$$

Note that:

$$\begin{aligned} E[Y_{0i} - Y_{1i} | Y_i = k, K_i^* = 0] &= \frac{1}{\mathcal{B}^*} \int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k) + \mathcal{B}^*} \{Q_{0|K^*=0}(u) - Q_{0|K^*=0}(u)\} du \\ &= \frac{1}{\mathcal{B}^*} \int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k) + \mathcal{B}^*} \{Q_{0|K^*=0}(u) - k\} du + \frac{1}{\mathcal{B}^*} \int_{F_{1|K^*=0}(k) - \mathcal{B}^*}^{F_{1|K^*=0}(k)} \{k - Q_{1|K^*=0}(v)\} dv \end{aligned}$$

where  $\mathcal{B}^* := P(h_{it} = k | K^* = 0)$ . A lower bound for  $E[Y_{0i} - Y_{1i} | Y_i = k, K_i^* = 0]$  is thus:

$$\begin{aligned} &\frac{F_{0|K^*=0}(k)}{f_{0|K^*=0}(k) \cdot \mathcal{B}^*} \int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k) + \mathcal{B}^*} \ln \left( \frac{u}{F_{0|K^*=0}(k)} \right) du + \frac{1 - F_{1|K^*=0}(k)}{f_{1|K^*=0}(k) \cdot \mathcal{B}^*} \int_{F_{1|K^*=0}(k) - \mathcal{B}^*}^{F_{1|K^*=0}(k)} \ln \left( \frac{1 - v}{1 - F_{1|K^*=0}(k)} \right) dv \\ &= g(F_{0|K^*=0}(k), f_{0|K^*=0}(k), \mathcal{B}^*) + h(F_{1|K^*=0}(k), f_{1|K^*=0}(k), \mathcal{B}^*) \end{aligned}$$

where

$$\begin{aligned} g(a, b, x) &:= \frac{a}{bx} \int_a^{a+x} \ln \left( \frac{u}{a} \right) du = \frac{a^2}{bx} \int_1^{1+\frac{x}{a}} \ln(u) du \\ &= \frac{a^2}{bx} \{u \ln(u) - u\} \Big|_1^{1+\frac{x}{a}} = \frac{a^2}{bx} \left\{ \left(1 + \frac{x}{a}\right) \ln \left(1 + \frac{x}{a}\right) - \frac{x}{a} \right\} \\ &= \frac{a}{bx} (a + x) \ln \left(1 + \frac{x}{a}\right) - \frac{a}{b} \end{aligned}$$

and

$$h(a, b, x) := \frac{1 - a}{bx} \int_{a-x}^a \ln \left( \frac{1 - v}{1 - a} \right) dv = \frac{(1 - a)^2}{bx} \int_1^{1+\frac{x}{1-a}} \ln(u) du = g(1 - a, b, x)$$

Similarly, an upper bound is:

$$\begin{aligned} &-\frac{1 - F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)(\mathcal{B}^*)} \int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k) + \mathcal{B}^*} \ln \left( \frac{1 - u}{1 - F_{0|K^*=0}(k)} \right) du \\ &\quad - \frac{F_{1|K^*=0}(k)}{f_{1|K^*=0}(k)(\mathcal{B}^*)} \int_{F_{1|K^*=0}(k) - \mathcal{B}^*}^{F_{1|K^*=0}(k)} \ln \left( \frac{v}{F_{1|K^*=0}(k)} \right) dv \\ &= \tilde{g}(F_{0|K^*=0}(k), f_{0|K^*=0}(k), \mathcal{B}^*) + \tilde{h}(F_{1|K^*=0}(k), f_{1|K^*=0}(k), \mathcal{B}^*) \end{aligned}$$

where

$$\begin{aligned}
\tilde{g}(a, b, x) &:= -\frac{1-a}{bx} \int_a^{a+x} \ln\left(\frac{1-u}{1-a}\right) du = -\frac{(1-a)^2}{bx} \int_{1-\frac{x}{1-a}}^1 \ln(u) du \\
&= \frac{(1-a)^2}{bx} \{u - u \ln(u)\} \Big|_{1-\frac{x}{1-a}}^1 = \frac{1-a}{b} + \frac{1-a}{bx} (1-a-x) \ln\left(1 - \frac{x}{1-a}\right) \\
&= -g(1-a, b, -x)
\end{aligned}$$

and

$$\tilde{h}(a, b, x) := -\frac{a}{bx} \int_{a-x}^a \ln\left(\frac{v}{a}\right) dv = -\frac{a^2}{bx} \int_{1-\frac{x}{a}}^1 \ln(u) du = \tilde{g}(1-a, b, x) = -g(a, b, -x)$$

Given  $p$ , we relate the  $K^* = 0$  conditional quantities to their unconditional analogues:

$$\begin{aligned}
F_{0|K^*=0}(k) &= \frac{F_0(k) - p}{1-p} \quad \text{and} \quad F_{1|K^*=0}(k) = \frac{F_1(k) - p}{1-p} \quad \text{and} \quad \mathcal{B}^* = \frac{\mathcal{B} - p}{1-p} \\
f_{0|K^*=0}(k) &= \frac{f_0(k)}{1-p} \quad \text{and} \quad f_{1|K^*=0}(k) = \frac{f_1(k)}{1-p}
\end{aligned}$$

Let  $F(h) = P(h_{it} \leq h)$  be the CDF of the data, and define  $f(h) = \frac{d}{dh} P(h_{it} \leq h)$  for  $h \neq k$ . By Lemma 1 and the BLC assumption, the above quantities are related to observables as:

$$F_0(k) = \lim_{h \uparrow k} F(h) + p, \quad F_1(k) = F(k), \quad f_0(k) = \lim_{h \uparrow k} f(h), \quad \text{and} \quad f_1(k) = \lim_{h \downarrow k} f(h)$$

As shown by Dümbgen et al. (2017), BLC implies the existence of a continuous density function, which assures that the required density limits exist, and delivers Item 1. of Theorem 1.

To obtain the final result, note that the function  $g(a, b, x)$  is homogeneous of degree zero. Thus  $\Delta_k^* \in [\Delta_k^L, \Delta_k^U :]$ , where

$$\Delta_k^L := g(F_-(k), f_-(k), \mathcal{B} - p) + g(1 - F(k), f_+(k), \mathcal{B} - p)$$

and

$$\Delta_k^U := -g(1 - p - F_-(k), f_-(k), p - \mathcal{B}) - g(F(k) - p, f_+(k), p - \mathcal{B})$$

where  $-$  and  $+$  subscripts denote left and right limits. The bounds are sharp as CHOICE, CONVEX and RANK imply no further restrictions on the potential outcome distributions.



## B.4 Proof of Theorem 2

This proof follows the notation of Appendix A. Throughout this proof we let  $Y_i(\rho, k) = Y_i(\rho)$ , given Assumption SEPARABLE. By Lemmas 2 and 3 the effect of changing  $k$  on bunching is:

$$\begin{aligned}\partial_k \{\mathcal{B} - p(k)\} &= -\frac{\partial}{\partial k} \int_{\rho_0}^{\rho_1} f_\rho(k) \mathbb{E} \left[ \frac{Y_i(\rho)}{d\rho} \middle| Y_i(\rho) = k \right] d\rho \\ &= -\int_{\rho_0}^{\rho_1} \frac{\partial}{\partial k} \left\{ f_\rho(k) \mathbb{E} \left[ \frac{Y_i(\rho)}{d\rho} \middle| Y_i(\rho) = k \right] \right\} d\rho \\ &= \int_{\rho_0}^{\rho_1} \partial_\rho f_\rho(k) d\rho = f_1(k) - f_0(k)\end{aligned}$$

Turning now to the total effect on average hours.

$$\begin{aligned}\partial_k E[Y_i^{[k, \rho_1]}] &= \partial_k \{P(Y_i(\rho_0) < k) \mathbb{E}[Y_i(\rho_0) | Y_i(\rho_0) < k]\} + k \partial_k (\mathcal{B}^{[k, \rho_1]} - p(k)) + \mathcal{B}^{[k, \rho_1]} - p(k) \\ &\quad + \partial_k \{P(Y_i(\rho_1) > k) \mathbb{E}[Y_i(\rho_1) | Y_i(\rho_1) > k]\} \\ &= \partial_k \int_{-\infty}^k y \cdot f_{\rho_0}(y) \cdot dy + k (f_0(k) - f_1(k)) + \mathcal{B}^{[k, \rho_1]} - p(k) + \partial_k \int_k^\infty y \cdot f_{\rho_1}(y) \cdot dy \\ &= \cancel{k f_0(k)} + k (f_1(k) - \cancel{f_0(k)}) + \mathcal{B}^{[k, \rho_1]} - p(k) - \cancel{k f_1(k)}\end{aligned}$$

Meanwhile:

$$\begin{aligned}\partial_{\rho_1} E[Y_i^{[k, \rho_1]}] &= k \partial_{\rho_1} \mathcal{B}^{[k, \rho_1]} + \partial_{\rho_1} \{P(Y_i(\rho_1) > k) \mathbb{E}[Y_i(\rho_1) | Y_i(\rho_1) > k]\} \\ &= k \partial_{\rho_1} \mathcal{B}^{[k, \rho_1]} + \int_k^\infty y \cdot \partial_{\rho_1} f_{\rho_1}(y) \cdot dy \\ &= -k f_{\rho_1}(k) \mathbb{E} \left[ \frac{Y_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = k \right] - \int_k^\infty y \cdot \partial_y \left\{ f_{\rho_1}(y) \mathbb{E} \left[ \frac{dY_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = y \right] \right\} dy \\ &= \cancel{-k f_{\rho_1}(k) \mathbb{E} \left[ \frac{Y_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = k \right]} + \cancel{y f_{\rho_1}(y) \mathbb{E} \left[ \frac{dY_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = y \right]} \Big|_k^\infty \\ &\quad - \int_k^\infty f_{\rho_1}(y) \mathbb{E} \left[ \frac{dY_i(\rho_1)}{d\rho} \middle| Y_i(\rho_1) = y \right] dy\end{aligned}$$

where I have used Lemma 2 with the Leibniz rule (establishing Item 3 in Theorem 2) as well as Lemma 3 in the third step, and then integration by parts along with the boundary condition that  $\lim_{y \rightarrow \infty} y \cdot f_{\rho_1}(y) = 0$ , implied by Assumption SMOOTH.

## C Incorporating workers that set their own hours

This section considers the robustness of the empirical strategy from Section 4 to a case where some workers are able to choose their own hours. In this case, a simple extension of the model leads to the bounds on the buncher ATE remaining valid, but it is only directly informative about the effects

of the FLSA among workers who have their hours chosen by the firm. In this section I follow the notation from the main text where  $h_{it}$  indicate the hours of worker  $i$  in week  $t$ .

Suppose that some workers are able to choose their hours each week without restriction (“worker-choosers”), and that for the remaining workers (“firm-choosers”) their employers set their hours. In general we can allow who chooses hours for a given worker to depend on the period, so let  $W_{it} = 1$  indicate that  $i$  is a worker-chooser in period  $t$ . Additionally, we continue to allow counterfactual bunchers for whom counterfactual hours satisfy  $h_{0it} = h_{1it} = 40$ , regardless of who chooses them. I replace Assumption CONVEX from Section 4 to allow agents to *either* dislike pay (firm-choosers), or like pay (worker-choosers):

**Assumption CONVEX\* (convex preferences, monotonic in either direction).** *For each  $i, t$  and function  $B(\mathbf{x})$ , choice is  $(c_{Bi}, \mathbf{x}_{Bi}) = \operatorname{argmax}_{c, \mathbf{x}} \{u_i(c, \mathbf{x}) : c \geq B(\mathbf{x})\}$  where  $u_i(c, \mathbf{x})$  is continuous and strictly quasi-concave in  $(c, \mathbf{x})$ , and*

- *strictly increasing in  $c$ , if  $W_{it} = 1$*
- *strictly decreasing in  $c$ , if  $W_{it} = 0$*

*Note:* This setup is general enough to also allow a stylized bargaining-inspired model in which choices maximize a weighted sum of quasilinear worker and firm utilities. For example, suppose that for any pay schedule  $B(h)$ :

$$h = \operatorname{argmax}_h \beta (f(h) - c) + (1 - \beta)(c - \nu(h)) \quad \text{with} \quad c = B(h) \quad (\text{C.6})$$

where  $f(h) - c$  is firm profits with concave production  $f$ ,  $c - \nu(h)$  is worker utility with a convex disutility of labor  $\nu(h)$ , and  $\beta \in [0, 1]$  governs the weight of each party in the negotiation (this corresponds to Nash bargaining in which outside options are strictly inferior to all  $h$  for both parties, and utility is log-linear in  $c$ ). Rearranging the maximand of Equation (C.6) as  $(1 - 2\beta)c + \{\beta f(h) - (1 - \beta)\nu(h)\}$ , we can observe that this setting delivers outcomes as-if chosen by a single agent with quasi-concave preferences, as  $\beta f(h) - (1 - \beta)\nu(h)$  is concave. For Assumption CONVEX from Section 4 to hold with the assumed direction of monotonicity in costs  $c$ , we would require that  $\beta > 1/2$  for all worker-firm pairs: informally, that firms have more say than workers do in determining hours. However CONVEX\* holds regardless of the distribution of  $\beta$  over worker-firm pairs. If  $\beta_{it} < 1/2$ , paycheck  $it$  will look exactly like a worker-chooser, and if  $\beta_{it} > 1/2$  paycheck  $it$  will look exactly like a firm-chooser.

In the generalized model of CONVEX\*, bunching is prima-facie evidence that firm-choosers exist, because there is no prediction of bunching among worker-choosers provided that potential

outcomes are continuously distributed (by contrast,  $k$  is a “hole” in the worker-chooser hours distribution). Indeed under regularity conditions all of the data local to 40 are from firm-choosers (and counterfactual bunchers). To make this claim precise, we assume that for worker-choosers hours are the only margin of response (i.e. their utility depends on  $\mathbf{x}$  only through  $y(\mathbf{x})$ ), and let  $IC_{0it}(y)$  and  $IC_{1it}(y)$  be the worker’s indifference curves passing through  $h_{0it}$  and  $h_{1it}$ , respectively. I assume these indifference curves are twice Lipschitz differentiable, with  $M_{it} := \sup_y \max\{|IC''_{0it}(y)|, |IC''_{1it}(y)|\}$ , where the supremum is taken over the support of hours, and  $IC''$  indicates second derivatives.

**Proposition 3.** *Suppose that the joint distribution of  $h_{0it}$  and  $h_{1it}$  admits a continuous density conditional on  $K_{it}^* = 0$ , and that for any worker-chooser  $IC_{0it}$  and  $IC_{1it}$  are differentiable with  $M_{it}/w_{it}$  having bounded support. Then, under CHOICE and CONVEX\*:*

- $P(h_{it} = k \text{ and } K_{it}^* = 0) = P(h_{1it} \leq k \leq h_{0it} \text{ and } K_{it}^* = 0 \text{ and } W_{it} = 0)$
- $\lim_{h \uparrow k} f(h) = P(W_{it} = 0) \lim_{h \uparrow k} f_{0|W=0}(h)$
- $\lim_{h \downarrow k} f(h) = P(W_{it} = 0) \lim_{h \downarrow k} f_{1|W=0}(h)$

*Proof.* See Supplemental Material. □

The first bullet of Proposition 3 says that all active bunchers are also firm-choosers, and have potential outcomes that straddle the kink. The second and third bullets state that the density of the data as hours approach 40 from either direction is composed only of worker-choosers. This result on density limits requires the stated regularity condition, which prevents worker indifference curves from becoming too close to themselves featuring a kink (plus a requirement that straight-time wages  $w_{it}$  be bounded away from zero).

Given the first item in Proposition 3, the buncher ATE introduced in Section 4 only includes firm-choosers:

$$\mathbb{E}[h_{0it} - h_{1it} | h_{it} = 40, K_{it}^* = 0] = \mathbb{E}[h_{0it} - h_{1it} | h_{it} = 40, K_{it}^* = 0, W_{it} = 0]$$

Accordingly, I assume rank invariance among the firm-chooser population only:

**Assumption RANK\*** (near rank invariance and counterfactual bunchers). *The following are true:*

- (a)  $P(h_{0it} = k) = P(h_{1it} = k) = p$
- (b)  $Y = k$  iff  $h_0 \in [k, k + \Delta_0^*]$  and  $W = 0$  iff  $h_1 \in [k - \Delta_1^*, k]$  and  $W = 0$ , for some  $\Delta_0^*, \Delta_1^*$

where  $p$  continues to denote  $P(K_{it}^* = 1)$ .

We may now state a version of Theorem 2 that conditions all quantities on  $W = 0$ , provided that we assume bi-log concavity of  $h_0$  and  $h_1$  conditional on  $W = 0$  and  $K = 0$ .

**Theorem 1\* (bi-log-concavity bounds on the buncher ATE, with worker-choosers).** *Assume CHOICE, CONVEX\* and RANK\* hold. If both  $h_{0it}$  and  $h_{1it}$  are bi-log concave conditional on the event ( $W_{it} = 0$  and  $K_{it}^* = 0$ ), then:*

$$\mathbb{E}[h_{0it} - h_{1it} | h_{it} = k, K_{it}^* = 0] \in [\Delta_k^L, \Delta_k^U]$$

where

$$\Delta_k^L = g(F_{0|W=0, K^*=0}(k), f_{0|W=0, K^*=0}(k), \mathcal{B}^*) + g(1 - F_{1|W=0, K^*=0}(k), f_{1|W=0, K^*=0}(k), \mathcal{B}^*)$$

and

$$\Delta_k^U = -g(1 - F_{0|W=0, K^*=0}(k), f_{0|W=0, K^*=0}(k), -\mathcal{B}^*) - g(F_{1|W=0, K^*=0}(k), f_{1|W=0, K^*=0}(k), -\mathcal{B}^*)$$

where  $\mathcal{B}^* = P(h_{it} = k | W_{it} = 0, K_{it}^* = 0)$  and

$$g(a, b, x) = \frac{a}{bx} (a + x) \ln \left( 1 + \frac{x}{a} \right) - \frac{a}{b}$$

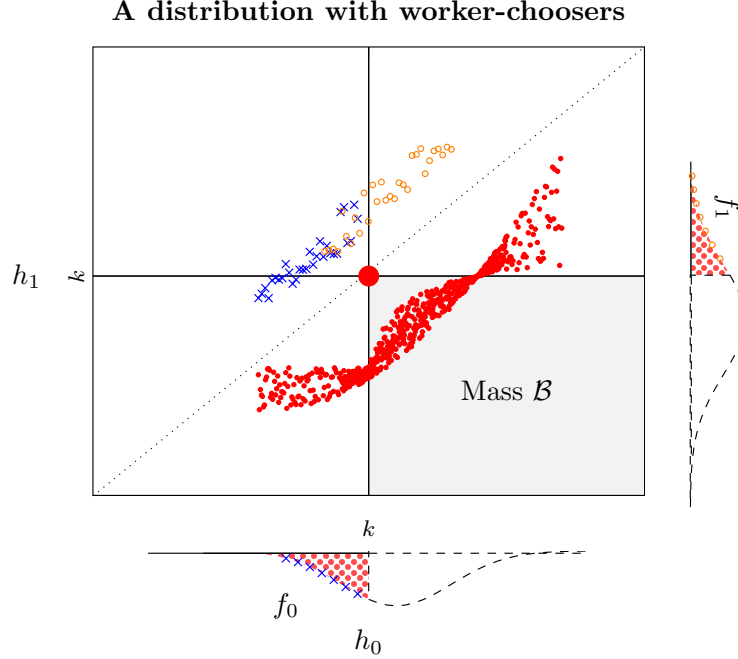
The bounds are sharp.

*Proof.* See Supplemental Appendix. □

Theorem 1\* does not immediately yield identification of the buncher-ATE bounds  $\Delta_k^L$  and  $\Delta_k^U$ , as we need to estimate each of the arguments to the function  $g$ . Using that the function  $g$  is homogenous of degree one, the bounds can be rewritten in terms of  $p$ , the identified quantities  $\mathcal{B}$ ,  $P(W_{it} = 0) \lim_{y \uparrow k} f_{0|W=0}(y)$  and  $P(W_{it} = 0) \lim_{y \uparrow k} f_{1|W=0}(y)$ , as well as the two probabilities  $P(h_{it} < 40 \text{ and } W_{it} = 0)$  and  $P(h_{it} > 40 \text{ and } W_{it} = 0)$  (see proof for details).

Figure C.4 depicts an example of a joint distribution of  $(h_0, h_1)$  that includes worker-choosers and satisfies Assumption RANK\*. The x-axis is  $h_0$ , and the y-axis is  $h_1$ , with the solid lines indicating 40 hours and the dotted diagonal line depicting  $h_1 = h_0$ . The dots show a hypothetical joint-distribution of the potential outcomes, with the (red) cloud south of the 45-degree line being firm-choosers, and the (blue and orange) cloud above being worker-choosers. Blue x's indicate worker-choosers who choose their value of  $h_0$ , while orange circles indicate worker-choosers who choose their value of  $h_1$ . The red dot at (40, 40) represents a mass of counterfactual bunchers.

Observed to the econometrician is the point mass at 40 as well as the truncated marginal distributions depicted at the bottom and the right of the figure, respectively. The observable



**FIGURE C.4:** The joint distribution of  $(h_{0it}, h_{1it})$ , for a distribution including worker-choosers and satisfying assumption RANK\*, cf. Figure 5. See text for description.

$P(h_{it} \leq h)$  for  $h < 40$  doesn't exactly identify  $P(h_{0it} \leq h)$  because some green x's are missing – these are worker-choosers for whom  $h_1 > 40 > h_0$  and choose to work overtime at their  $h_1$  value. Thus they show up in the data at  $h > 40$  even though they have  $h_0 < 40$ . Similarly, some blue circles are missing from the data above 40 – these are worker-choosers for whom  $h_1 > 40 > h_0$  and choose to work their  $h_0$  value, not working overtime. The probabilities  $P(h_{it} < 40 \text{ and } W_{it} = 0)$  and  $P(h_{it} > 40 \text{ and } W_{it} = 0)$  can thus only be estimated with some error, with the size of the error depending on the mass of worker-choosers in the northwest quadrant of Figure C.4. However, this has little impact on the results.<sup>50</sup>

Two further caveats of Theorem 1\* are worth mentioning here. First, an evaluation of the FLSA would ideally account for worker-choosers (who are working longer hours as a result of the policy) when averaging treatment effects. However, the proportion of worker-choosers and the size of their hours increases are not identified. Using the buncher ATE to estimate the overall ex-post effect of the FLSA – as described in Section 4.4 – may overstate its overall average net hours reduction. Secondly, note that we can no longer directly verify the bi-log concavity assumption of  $h_0$  for  $h < k$ , and of  $h_1$  for  $h > k$ , by looking at the data. The reason is that the observed data is a mixture of the firm-chooser and worker-chooser distributions, while our BLC assumption regards the subgroup of firm-choosers. If the proportion of worker-choosers is small, then these caveats

<sup>50</sup>The components of the bounds  $\Delta_k^L = L0 + L1$  and  $\Delta_k^U = -U0 - U1$  are not very sensitive to the values of the CDF inputs  $F_{0|W=0, K^*=0}(k)$  and  $F_{1|W=0, K^*=0}(k)$ , as can be verified numerically (details available upon request). Intuitively,  $\Delta_k^L$  and  $\Delta_k^U$  mostly depend on the density estimates and the size of the bunching mass.

should have only a minor impact on the interpretation of the results. The first problem is difficult to avoid: estimating the overall effect of the FLSA based on a subset of firm-choosers is inevitably going to miss the fact that overtime pay increases hours for some workers. However, the survey evidence mentioned in Section 2 suggests that the set of such workers is relatively small.

## D Interdependencies among hours within the firm

In this section I consider the impact that interdependencies among the hours of different units may have on the estimates, reflected in the third term of Equation (8) from Section 4.4. I develop some structure to guide our intuition of this term, and then present some empirical evidence that it is likely to be small.

The basic issue is as follows: when a single firm chooses hours jointly among multiple units—either across different workers or across multiple weeks, or both—this term may be nonzero and contribute to the overall effect of the FLSA. This can be thought of as a violation of the stable unit treatment value assumption (SUTVA) in using the treatment effects  $\Delta_{it}$  to assess the average impact of the FLSA on hours, captured by the third term of Equation (8).

To simplify the notation, I'll assume that such SUTVA violations may occur across workers within a firm in a single week, suppressing the time index  $t$  and focusing on a single firm. As in Section 4.4 let  $\mathbf{h}_{-i}$  denote the vector of actual (observed) hours for all workers aside from  $i$  within  $i$ 's firm. These hours are chosen according to the kinked cost schedule introduced by the FLSA. Let  $h_{0i}(\cdot)$  denote the hours that the firm would choose for worker  $i$  if they had to pay  $i$ ' straight-wage  $w_i$  for all of  $i$ 's hours, as a function of the hours profile of the other workers in the firm (suppressing dependence on straight-wages in this section). Define  $h_{1i}(\cdot)$  analogously with  $1.5w_i$ . In this notation, the potential outcomes from Section 4 are  $h_{0i} = h_{0i}(\mathbf{h}_{-i})$  and  $h_{1i} = h_{1i}(\mathbf{h}_{-i})$ . As in Section 4.4 let  $(h_i^*, \mathbf{h}_{-i}^*)$  denote the hours profile that would occur absent the FLSA, so that the average ex-post effect of the FLSA is  $\mathbb{E}[h_i - h_i^*]$ .

For concreteness, let us suppose that hours are chosen to maximize profits with a joint-production function  $F(\mathbf{h})$ , where  $\mathbf{h}$  is a vector of the hours this week across all workers in the firm. We then have that  $(h_i, \mathbf{h}_{-i}) = \operatorname{argmax} \{F(\mathbf{h}) - \sum_j B_{kj}(h_j)\}$ , where the sum is across workers  $j$  and  $B_{kj}(h) := w_j h + .5w_j \mathbb{1}(h > 40)(h - 40)$ . Similarly  $(h_i^*, \mathbf{h}_{-i}^*) = \operatorname{argmax} \{F(\mathbf{h}) - \sum_j w_j h_j\}$ . Whether  $h_{0i}(\mathbf{h}_{-i})$  is smaller or larger than  $h_i^*$  (with a fixed set of employees) will depend upon whether  $i$ 's hours are complements or substitutes in production with those of each of their colleagues, and with what strength. It is natural to expect that either case might occur. Consider for example a production function in which workers are divided into groups  $\theta_1 \dots \theta_M$  corresponding

to different occupations, and:

$$F(\mathbf{h}) = \prod_{m=1}^M \left( \left( \sum_{i \in \theta_m} a_i \cdot h_i^{\rho_m} \right)^{1/\rho_m} \right)^{\alpha_m} \quad (\text{D.7})$$

where  $a_i$  is an individual productivity parameter for worker  $i$ . The hours of workers within an occupation enter as a CES aggregate with substitution parameter  $\rho_m$ , which then combine in a Cobb-Douglas form across occupations with exponents  $\alpha_m$ . The hours of two workers  $i$  and  $j$  belonging to different occupations are always complements in production, i.e.  $\partial_{h_i} F(\mathbf{h})$  is increasing in  $h_j$ . When  $i$  and  $j$  belong to the same occupation  $\theta_m$ , it can be shown that worker  $i$  and  $j$ 's hours are substitutes—i.e.  $\partial_{h_i} F(\mathbf{h})$  is *decreasing* in  $h_j$ —when  $\alpha_m \leq \rho_m$ .

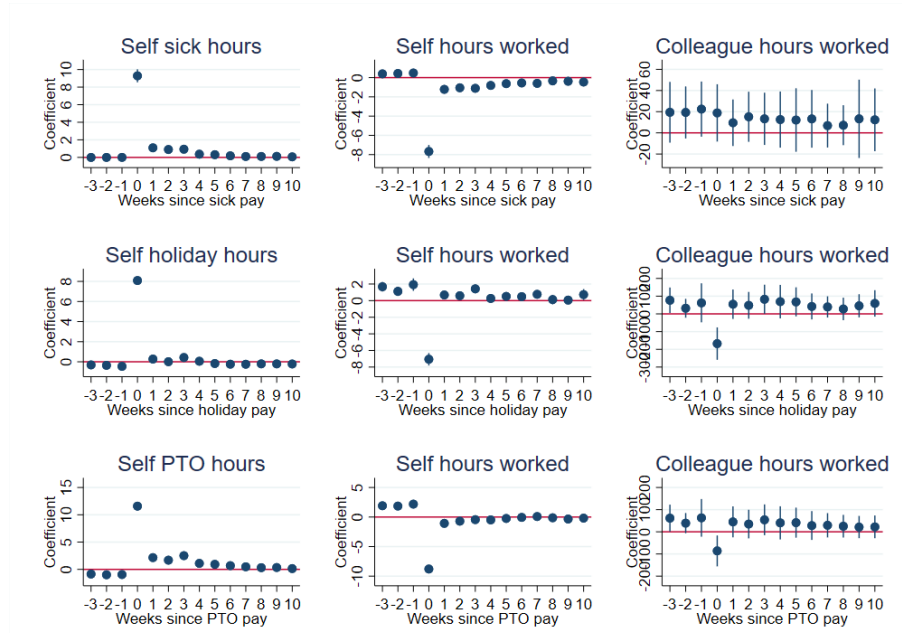
Thus both substitution and complementarity in hours can plausibly coexist within a firm, and it is difficult to sign theoretically the contribution of interdependencies on our parameter of interest  $\theta$  (c.f. Eq. (8)). Given that neither occupations nor tasks are observed in the data, it is also difficult to obtain direct evidence even with the aid of functional-form assumptions like Eq. (D.7). I therefore turn to an indirect empirical test of whether these effects are likely to play a significant role in  $\theta$ .

Figure D.5 shows that in weeks when a worker receives a positive number of sick-pay hours, their individual hours worked for that week decline by about 8 hours on average. Yet I fail to find evidence of a corresponding change in the hours of others in the same firm. This suggests that short term variation in the hours of a worker's colleagues does not tend to translate into contemporaneous changes in their own (for example, if the firm were dividing a fixed number of hours across workers).

Table D.1 shows another piece of evidence: that my overall effect estimates are similar between small, medium, and large firms. If firms were to compensate for overtime hours reductions by “giving” some hours to similar workers who would otherwise be working less than 40, for instance, then we would expect this to play a larger role in firms where there are a large number of substitutable workers—causing a bias that increases with firm size. I cannot reject that my strategy estimates the same parameter value across the three firm size categories, in my preferred specification of estimating  $p$  using variation in PTO.

	$p=0$		$p$ from PTO	
	Bunching	Effect of the kink	Net Bunching	Effect of the kink
Small firms	0.198 [0.189, 0.208]	[-1.525, -1.455] [-1.676, -1.299]	0.027 [0.023, 0.031]	[-0.231, -0.171] [-0.274, -0.139]
Medium firms	0.103 [0.095, 0.110]	[-1.123, -0.786] [-1.237, -0.710]	0.030 [0.025, 0.035]	[-0.337, -0.224] [-0.407, -0.178]
Large firms	0.050 [0.047, 0.054]	[-0.768, -0.468] [-0.861, -0.414]	0.024 [0.021, 0.028]	[-0.371, -0.224] [-0.444, -0.180]

**TABLE D.1:** Estimates of the ex-post effect of the kink by firm size. “Small” firms have between 1 and 25 workers in my estimation sample, “Medium” have 26 to 50, and “Large” have more than 50. Note that the estimated net bunching caused by the FLSA is similar across firm sizes (right), despite the raw bunching observed in the data differing by firm size category.



**FIGURE D.5:** Event study coefficients  $\beta_j$  and 95% confidence intervals across an instance of a worker receiving pay for non-work hours (either sick pay, holiday pay, or paid time off-‘PTO’). Equation is  $y_{it} = \mu_t + \lambda_i + \sum_{j=-3}^{10} \beta_j D_{it,j} + u_{it}$ , where  $D_{it,j} = 1$  if worker  $i$  in week  $t$  has a positive number of a given type of non-work hours  $j$  weeks ago (after a period of at least three weeks in which they did not),  $\lambda_i$  are worker fixed effects, and  $\mu_t$  are calendar week effects. Rows correspond to choices of the non-work pay type: either sick, holiday, PTO. Columns indicate choices of the outcome  $y_{it}$ . “Colleague hours worked” sums the hours of work in  $t$  across all workers other than  $i$  in  $i$ ’s firm. The timing of holiday and PTO hours appears to be correlated across workers, leading to a decrease in the working hours of  $i$ ’s colleagues in weeks in which  $i$  takes either holiday or PTO pay (center-right and bottom-right graphs). However I cannot reject that colleague work hours are unrelated to an instance of sick pay: before, during and after it occurs (top-right). Since  $i$ ’s hours of work reduce by about 8 hours on average during an instance of sick pay (top-center), this suggests that there is no contemporaneous reallocation of  $i$ ’s forgone work hours to their colleagues.



## E A simple model of wages and “typical” hours

Each firm faces a labor supply curve  $N(z, h)$ , indicating the labor force  $N$  it can maintain if it offers total compensation  $z$  to each of its workers, when they are each expected to work  $h$  hours per week. The firm chooses a pair  $(z^*, h^*)$  based on the cost-minimization problem:

$$\min_{z, h, K, N} N \cdot (z + \psi) + rK \quad \text{s.t.} \quad F(Ne(h), K) \geq Q \quad \text{and} \quad N \leq N(z, h) \quad (\text{E.8})$$

where the labor supply function is increasing in  $z$  while decreasing in  $h$ ,  $e(h)$  represents the “effective labor” from a single worker working  $h$  hours, and  $\psi$  represents non-wage costs per worker. The quantity  $\psi$  can include for example recruitment effort and training costs, administrative overhead and benefits that do not depend on  $h$ . Concavity of  $e(h)$  captures declining productivity at longer hours, for example from fatigue or morale effects. The function  $F$  maps total effective labor  $Ne(h)$  and capital into level of output or revenue that is required to meet a target  $Q$ , and  $r$  is the cost of capital. For simplicity, workers within a firm are here identical and all covered by the FLSA.

To understand the properties of the solution to Equation (E.8), let us examine two illustrative special cases.

### Special case 1: an exogenous competitive straight-time wage

Much of the literature on hours determination has taken the hourly wage as a fixed input to the choice of hours, and assumed that at that wage the firm can hire any number of workers, regardless of hours. This can be motivated as a special case of Equation (E.8) in which there is perfect competition on the straight-time wage, i.e.  $N(z, h) = \bar{N} \mathbb{1}(w_s(z, h) \geq w)$  for some large number  $\bar{N}$  and wage  $w$  exogenous to the firm. Then Equation (E.8) reduces to:

$$\min_{N, h, K} N \cdot (hw + \mathbb{1}(h > 40)(w/2)(h - 40) + \psi) + rK \quad \text{s.t.} \quad F(Ne(h), K) \geq Q \quad (\text{E.9})$$

By limiting the scope of labor supply effects in the firm’s decision, Equation (E.9) is well-suited to illustrating the competing forces that shape hours choice on the production side: namely the fixed costs  $\psi$  and the concavity of  $e(h)$ . Were  $\psi$  equal to zero with  $e(h)$  strictly concave globally, a firm solving Equation (E.9) would always find it cheaper to produce a given level of output with more workers working less hours each. On the other hand, were  $\psi$  positive and  $e$  weakly convex, it would always be cheapest to hire a single worker to work all of the firm’s hours. In general, fixed costs and declining hours productivity introduce a tradeoff that leads to an interior solution for hours.<sup>51</sup>

---

<sup>51</sup>In the fixed-wage special case, these two forces along with the wage are in fact sufficient to pin down hours, which do not depend on the production function  $F$  or the chosen output level  $Q$ . See e.g. Cahuc and Zylberberg (2004) for the case in which  $e(h)$  is iso-elastic.

Equation (E.9) introduces a kink into the firm's costs as a function of hours, much as short-run wage rigidity does in my dynamic analysis. However, the assumption that the firm can demand any number of hours at a set straight-time wage rate is harder to defend when thinking about firms long-run expectations, a point emphasized by Lewis (1969). Equilibrium considerations will also tend to run against the independence of hourly wages and hours - a mechanism explored in Supplemental Appendix 1.

## Special case 2: iso-elastic functional forms

By placing some functional form restrictions on Equation (E.8), we can obtain a closed-form expression for  $(z^*, h^*)$ . In particular, when labor supply and  $e(h)$  are iso-elastic, production is separable between capital and labor and linear in the latter, and firms set the output target  $Q$  to maximize profits, Proposition 4 characterizes the firm's choice of earnings and hours:

**Proposition 4.** *When i)  $e(h) = e_0 h^\eta$  and  $N(z, h) = N_0 z^{\beta_z} h^{\beta_h}$ ; ii)  $F(L, K) = L + \phi(K)$  for some function  $\phi$ ; and iii)  $Q$  is chosen to maximize profits, the  $(z^*, h^*)$  that solve Equation (E.8) are:*

$$h^* = \left[ \frac{\psi}{e_0} \cdot \frac{\beta}{\beta - \eta} \right]^{1/\eta} \quad \text{and} \quad z^* = \psi \cdot \frac{\beta_z}{\beta_z + 1} \frac{\eta}{\beta - \eta}$$

where  $\beta := \frac{|\beta_h|}{\beta_z + 1}$ , provided that  $\psi > 0$ ,  $\eta \in (0, \beta)$ ,  $\beta_h < 0$  and  $\beta_z > 0$ . Hours and compensation are both decreasing in  $|\beta_h|$  and increasing in  $\beta_z$ .

*Proof.* See Supplemental Material. □

The proposition shows that the hours chosen depend on labor supply via  $\beta = \frac{|\beta_h|}{1 + \beta_z}$ , which gauges how elastic labor supply is with respect to hours compared with earnings. The more sensitive labor supply is to a marginal increase in hours as compared with compensation, the higher  $\beta$  will be and lower the optimal number of hours. The proof of Proposition 4 also shows that unlike Special case 1 of perfect competition on the straight-time wage, when  $N(z, h)$  is differentiable the general model can support an interior solution for hours even without fixed costs  $\psi = 0$ .

*Note:* Broadly speaking, the function  $N(z, h)$  might be viewed as an equilibrium object that reflects both worker preferences over income and leisure and the competitive environment for labor. Thus it is conceivable that equilibrium forces lead to a labor supply function like that of the fixed-wage model, in which the the FLSA has an effect on the hours set at hiring. In Supplemental Appendix 1, I show that the prediction of the fixed-job model that the FLSA has little to no effect on  $h^*$  or  $z^*$  is robust to embedding Equation (E.8) into an extension of the Burdett and Mortensen (1998)

model of equilibrium with on-the-job search.<sup>52</sup> In the context of the search model, the only effect of the overtime rule on the distribution of  $h^*$  is mediated through the minimum wage, which rules out some of the  $(z^*, h^*)$  pairs that would occur in the unregulated equilibrium. In a numerical calibration, this effect is quite small, suggesting that equilibrium effects play only a minor role in how the FLSA overtime rule impacts anticipated hours or straight-time wages.

## F Additional empirical information and results

### F.1 Sample restrictions

As discussed in Section 3, I keep paychecks from workers who are paid on a weekly basis, and condition on paychecks that contain a record of positive hours for work, vacation, holidays, or sick leave, totaling fewer than 80 hours in a week.<sup>53</sup> I also drop observations from California, which has a daily overtime rule that is binding for a significant number of workers, and could confound the effects of the weekly FLSA rule.

Further, I focus on hourly workers. While the data include a field for the employer to input a salary, there is no guarantee that employers actually use this feature in the payroll software. Therefore, I use a combination of sampling restrictions to ensure I remove all non-hourly workers from the sample. First, I drop workers that ever have a salary on file with the payroll system. Second, I only keep workers at firms for whom *some* workers have a salary on file, reflecting an assumption that employers either don't use the feature at all or use it for all of their salaried employees. I also drop paychecks from workers for whom hours are recorded as 40 in every week in the sample,<sup>54</sup> as it is possible that these workers are simply coded as working 40 hours despite being paid on a salary basis. I also drop workers who never receive overtime pay.

### F.2 A test of the Trejo (1991) model of straight-time wage adjustment

Another way to assess the role of the wage rigidity reported in Table 2 is to test directly whether straight-time wages and hours are plausibly related *at the weekly level* according to Equation (1). We can do this using the wage and hours reported on each paycheck, and given the kink in Eq. (1) making only weak differentiability assumptions on unobservables for identification.

---

<sup>52</sup>This remains true even in the perfectly competitive limit of the model, the basic reason being that workers choose to accept jobs on the basis of their known total earnings  $z^*$ , rather than the straight-time wage.

<sup>53</sup>This restriction removes about 2% of the sample after the other restrictions. While a genuine 80 hour workweek is possible, I consider these observations to likely correspond to two weeks of work despite the worker's pay frequency being coded as weekly.

<sup>54</sup>For the purposes of this restriction, I count the "40 hours" event as occurring when either hours worked or hours paid is equal to 40.

Suppose that the wages for some workers are actively adjusted to the hours they work according to Equation (1), in order to target some total earnings  $z_{it}$ . Denote the corresponding observational units  $it$  by a latent variable  $A_{it} = 1$ . Units with  $A_{it} = 1$  may be workers with almost no variation in their schedules, for whom their wages were set according to Eq. (1) at hiring, or their wages may be dynamic and adjust to week-by-week variation in their hours. Let  $A_{it} = 0$  denote units for whom the worker's wage is determined in some other way.

Let  $q(h) = P(A_{it} = 1|h_{it} = h)$  denote the proportion of these two groups at various points in the hours distribution. An extreme version of the fixed-job model of Trejo (1991) for example, would have  $q(h) = 1$  for all  $h$ .

By the law of iterated expectations and some algebra we have that:

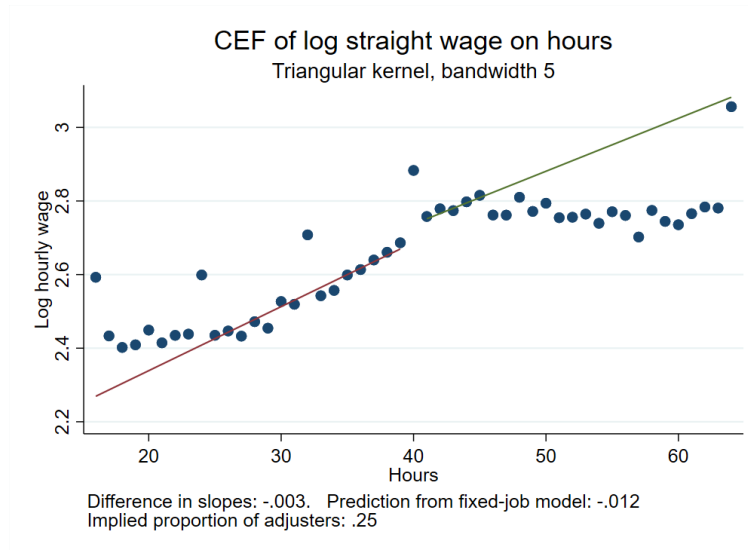
$$\begin{aligned} \mathbb{E} [\ln w_{it}|h_{it} = h] &= q(h) \{ \mathbb{E} [z_{it}|h_{it} = h, A_{it} = 1] - \ln(h + 0.5(h - 40)\mathbb{1}(h \geq 40)) \} \\ &\quad - (1 - q(h)) \mathbb{E} [\ln w_{it}|h_{it} = h, A_{it} = 0] \end{aligned}$$

The middle term above introduces a kink in the conditional expectation of log wages with respect to hours. If we assume that  $\mathbb{E} [\ln z_{it}|h_{it} = h, A_{it} = 1]$ ,  $\mathbb{E} [\ln w_{it}|h_{it} = h, A_{it} = 0]$  and  $q(h)$  are all continuously differentiable in  $h$ , then the magnitude of this kink identifies  $q(40)$ , the proportion of active wage responders local to  $h = 40$ :

$$\lim_{h \downarrow 40} \frac{d}{dh} \mathbb{E} [\ln w_{it}|h_{it} = h] - \lim_{h \uparrow 40} \frac{d}{dh} \mathbb{E} [\ln w_{it}|h_{it} = h] = -\frac{1}{2} \cdot \frac{q(40)}{40}$$

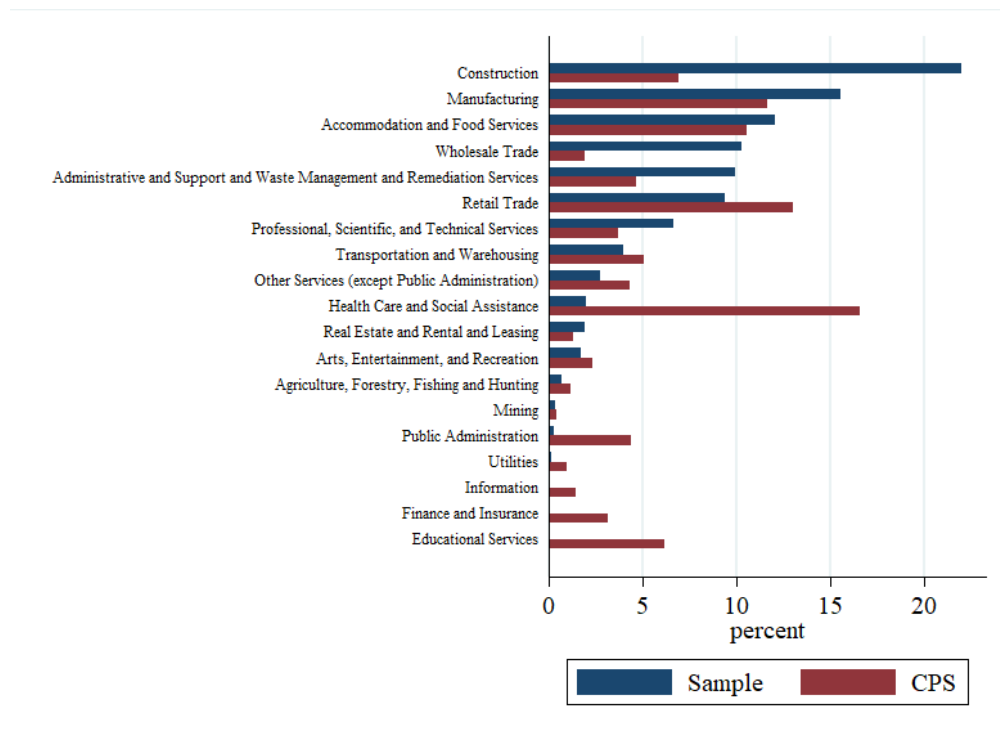
These continuous differentiability assumptions are reasonable, if wage setting according to Equation (1) is the only force introducing non-smoothness in the relationship between wages and hours. For instance, we assume that production technologies do not have any special features at 40 hours that would cause the distribution of target earnings levels  $z_{it}$  among the  $A_{it} = 1$  units to itself have a kink around  $h_{it} = 40$ .

Figure F.6 reports the results of fitting separate local linear functions to the CEF of log wages on either side of  $h = 40$ . We can reject the hypothesis that the fixed-job model applies to all employees at all times, near 40. However, the data appear to be consistent with a proportion  $q(40)$  of about 0.25 of all paychecks close to 40 hours reflecting an hours/wage relationship governed by Equation (1). This can be rationalized by straight-wages being updated intermittently to reflect expected or anticipated hours, which vary in practice quite a bit between pay periods.

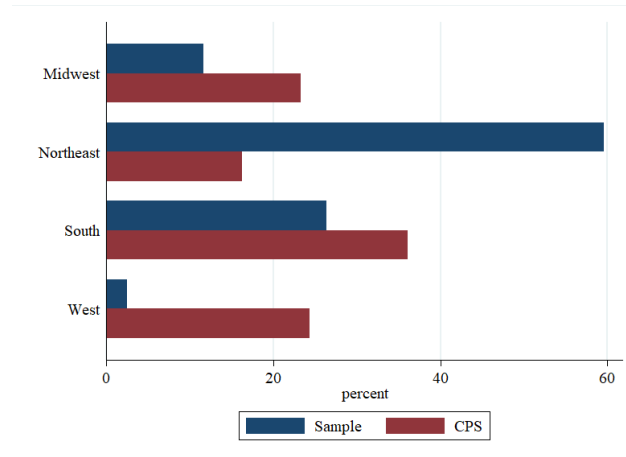


**FIGURE F.6:** A kinked-CEF test of the fixed-jobs model presented in Trejo (1991). Regression lines fit on each side with a uniform kernel within 25 hours of the 40.

### F.3 Further characteristics of the sample



**FIGURE F.7:** Industry distribution of estimation sample versus the Current Population Survey sample described in Section 3.



**FIGURE F.8:** Geographical distribution of estimation sample versus the Current Population Survey sample described in Section 3.

Industry	Avg. OT hours	OT % hours	OT % pay	Industry share
Accommodation and Food Services	2.37	0.06	0.11	0.08
Administrative and Support	5.69	0.13	0.18	0.08
Agriculture, Forestry, Fishing and Hunting	3.76	0.11	0.15	0.00
Arts, Entertainment, and Recreation	3.87	0.10	0.13	0.00
Construction	3.09	0.07	0.10	0.20
Educational Services	1.83	0.05	0.07	0.00
Finance and Insurance	0.31	0.00	0.01	0.00
Health Care and Social Assistance	4.59	0.12	0.12	0.02
Information	1.67	0.04	0.06	0.00
Manufacturing	3.37	0.08	0.11	0.18
Mining	2.26	0.07	0.12	0.00
Other Services	2.61	0.06	0.09	0.02
Professional, Scientific, and Technical Services	2.91	0.07	0.10	0.06
Public Administration	2.36	0.05	0.08	0.00
Real Estate and Rental and Leasing	2.85	0.07	0.09	0.02
Retail Trade	2.83	0.07	0.10	0.08
Transportation and Warehousing	5.24	0.12	0.17	0.04
Utilities	3.80	0.08	0.11	0.00
Wholesale Trade	5.15	0.11	0.14	0.10
Total Sample	3.55	0.08	0.12	0.98

**TABLE F.2:** Overtime prevalence by industry in the sample, including average number of OT hours per weekly paycheck, % OT hours among hours worked, % pay for hours work going to OT, and industry share of total hours in sample.

	(1)	(2)	(3)	(4)	(5)
	Work hours=40	OT hours	Total work hours	Work hours=40	OT hours
Tenure	0.000400 (0.95)	0.0515 (3.95)	0.0796 (3.31)		
Age	0.000690 (3.82)	0.00266 (0.74)	0.0250 (3.25)		
Female	0.0140 (2.08)	-1.322 (-9.07)	-1.943 (-6.08)		
Minimum wage worker	0.00121 (0.29)	-1.687 (-2.39)	-5.352 (-4.08)		
Firm just hired				-0.00572 (-2.95)	0.553 (5.78)
Date FE	Yes	Yes	Yes	Yes	Yes
Employer FE	Yes	Yes	Yes		
Worker FE				Yes	Yes
Observations	499619	499619	499619	628449	628449
R squared	0.229	0.264	0.260	0.387	0.515

*t* statistics in parentheses

**TABLE F.3:** Columns (1)-(3) regress hours-related outcome variables on worker characteristics, with fixed effects for the date and employer. Standard errors clustered by firm. Columns (4)-(5) show that bunching and overtime hours among incumbent workers are both responsive to new workers being hired within a firm, even controlling for worker and day fixed effects. “Firm just hired” indicates that at least one new worker appears in payroll at the firm this week, and the new workers are dropped from the regression. “Minimum wage worker” indicates that the worker’s straight-time wage is at or below the maximum minimum wage in their state of residence for the quarter. Tenure and age are measured in years, and age is missing for some workers.

	(1)	(2)	(3)
	Total work hours	Total work hours	Total work hours
R squared	0.366	0.499	0.626
Date FE		Yes	
Worker FE		Yes	Yes
Employer x date FE	Yes		Yes
Observations	621011	628449	620854

*t* statistics in parentheses

**TABLE F.4:** Decomposing variation in total hours. Worker fixed effects and employer by day fixed effects explain about 63% of the variation in total hours.

## F.4 Additional treatment effect estimates and figures

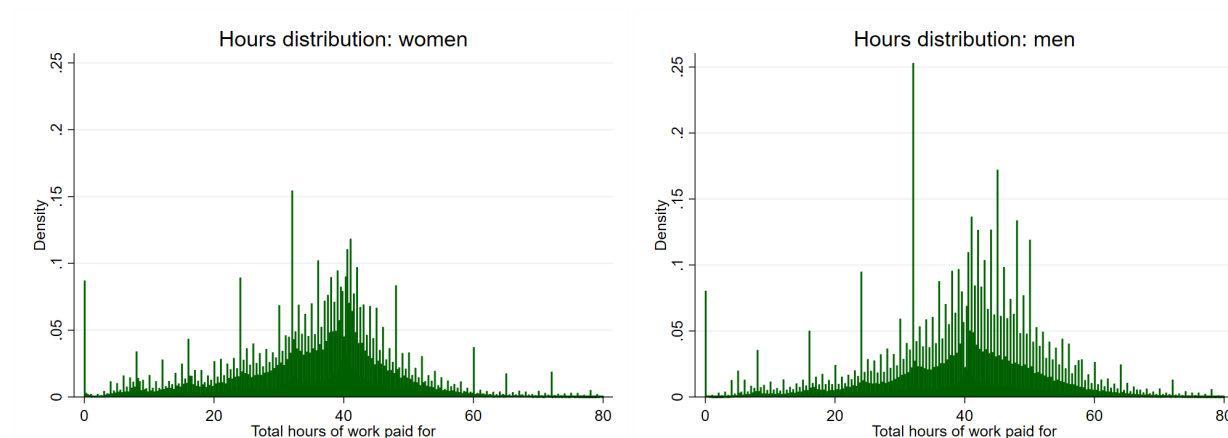
	$p=0$		$p$ from PTO	
	Bunching	Buncher ATE	Net Bunching	Buncher ATE
Accommodation and Food Services (N=69427)	0.036 [0.029, 0.044]	[0.937, 0.988] [0.734, 1.212]	0.036 [0.029, 0.044]	[0.937, 0.988] [0.734, 1.212]
Administrative and Support (N=49829)	0.062 [0.051, 0.074]	[1.625, 1.771] [1.313, 2.136]	0.009 [0.005, 0.013]	[0.251, 0.255] [0.143, 0.365]
Construction (N=136815)	0.139 [0.128, 0.149]	[2.759, 3.326] [2.341, 3.854]	0.029 [0.022, 0.035]	[0.612, 0.638] [0.442, 0.821]
Health Care and Social Assistance (N=13951)	0.051 [0.034, 0.069]	[1.412, 1.522] [0.570, 2.450]	0.005 [0.000, 0.010]	[0.146, 0.147] [-0.052, 0.348]
Manufacturing (N=112555)	0.137 [0.126, 0.148]	[2.098, 2.521] [1.894, 2.785]	0.018 [0.016, 0.021]	[0.307, 0.316] [0.255, 0.370]
Other Services (N=19263)	0.160 [0.132, 0.188]	[1.804, 2.240] [1.243, 2.996]	0.037 [0.024, 0.049]	[0.452, 0.478] [0.256, 0.693]
Professional, Scientific, Technical (N=47705)	0.136 [0.117, 0.155]	[2.281, 2.737] [1.862, 3.297]	0.010 [0.003, 0.016]	[0.178, 0.180] [0.060, 0.302]
Real Estate and Rental and Leasing (N=13498)	0.187 [0.141, 0.234]	[3.477, 4.478] [2.432, 6.053]	0.097 [0.060, 0.135]	[1.920, 2.215] [1.065, 3.316]
Retail Trade (N=56403)	0.129 [0.112, 0.146]	[3.694, 4.399] [2.447, 5.935]	0.032 [0.024, 0.040]	[0.969, 1.016] [0.550, 1.463]
Transportation and Warehousing (N=25926)	0.091 [0.070, 0.111]	[2.230, 2.530] [1.754, 3.127]	0.015 [0.009, 0.022]	[0.400, 0.409] [0.216, 0.602]
Wholesale Trade (N=66678)	0.126 [0.110, 0.141]	[2.751, 3.299] [2.321, 3.848]	0.046 [0.037, 0.055]	[1.068, 1.149] [0.765, 1.490]
All Industries (N=630217)	0.116 [0.112, 0.121]	[2.614, 3.054] [2.483, 3.217]	0.027 [0.024, 0.029]	[0.640, 0.666] [0.571, 0.740]

**TABLE F.5:** Estimates of the buncher ATE by industry, based on  $p = 0$  (left) or  $p$  estimated from paid time off (right). Estimates are reported only for industries having at least 10,000 observations. 95% bootstrap confidence intervals in gray, clustered by firm.



	$p=0$		$p$ from PTO	
	Bunching	Effect of the kink	Net Bunching	Effect of the kink
Accommodation and Food Services (N=69427)	0.036 [0.029, 0.044]	[-0.368, -0.248] [-0.450, -0.192]	0.036 [0.029, 0.044]	[-0.368, -0.248] [-0.450, -0.192]
Administrative and Support (N=49829)	0.062 [0.051, 0.074]	[-1.190, -0.681] [-1.424, -0.548]	0.009 [0.005, 0.013]	[-0.178, -0.101] [-0.256, -0.057]
Construction (N=136815)	0.139 [0.128, 0.149]	[-1.550, -1.121] [-1.771, -0.944]	0.029 [0.022, 0.035]	[-0.330, -0.219] [-0.422, -0.157]
Health Care and Social Assistance (N=13951)	0.051 [0.034, 0.069]	[-0.633, -0.320] [-1.020, -0.129]	0.005 [0.000, 0.010]	[-0.065, -0.030] [-0.155, 0.012]
Manufacturing (N=112555)	0.137 [0.126, 0.148]	[-1.167, -0.850] [-1.282, -0.766]	0.018 [0.016, 0.021]	[-0.162, -0.110] [-0.192, -0.090]
Other Services (N=19263)	0.160 [0.132, 0.188]	[-0.977, -0.811] [-1.300, -0.538]	0.037 [0.024, 0.049]	[-0.235, -0.176] [-0.345, -0.095]
Professional, Scientific, Technical (N=47705)	0.136 [0.117, 0.155]	[-1.192, -0.959] [-1.411, -0.767]	0.010 [0.003, 0.016]	[-0.090, -0.063] [-0.150, -0.021]
Real Estate and Rental and Leasing (N=13498)	0.187 [0.141, 0.234]	[-1.766, -1.466] [-2.303, -1.002]	0.097 [0.060, 0.135]	[-0.954, -0.725] [-1.378, -0.392]
Retail Trade (N=56403)	0.129 [0.112, 0.146]	[-1.685, -1.342] [-2.274, -0.908]	0.032 [0.024, 0.040]	[-0.434, -0.308] [-0.626, -0.175]
Transportation and Warehousing (N=25926)	0.091 [0.070, 0.111]	[-1.590, -0.998] [-1.935, -0.783]	0.015 [0.009, 0.022]	[-0.274, -0.166] [-0.406, -0.086]
Wholesale Trade (N=66678)	0.126 [0.110, 0.141]	[-2.122, -1.297] [-2.474, -1.088]	0.046 [0.037, 0.055]	[-0.776, -0.476] [-1.016, -0.333]
All Industries (N=630217)	0.116 [0.112, 0.121]	[-1.466, -1.026] [-1.542, -0.972]	0.027 [0.024, 0.029]	[-0.347, -0.227] [-0.386, -0.202]

**TABLE F.6:** Estimates of the hours effect of the FLSA by industry, based on  $p = 0$  (left) or  $p$  estimated from paid time off (right). Estimates are reported only for industries having at least 10,000 observations. 95% bootstrap confidence intervals in gray, clustered by firm. In the case of Accommodation and Food Services,  $P(h_{it} = 40 | \eta_{it} > 0) > \mathcal{B}$ , so I take the PTO-based estimate to be  $p = 0$ .



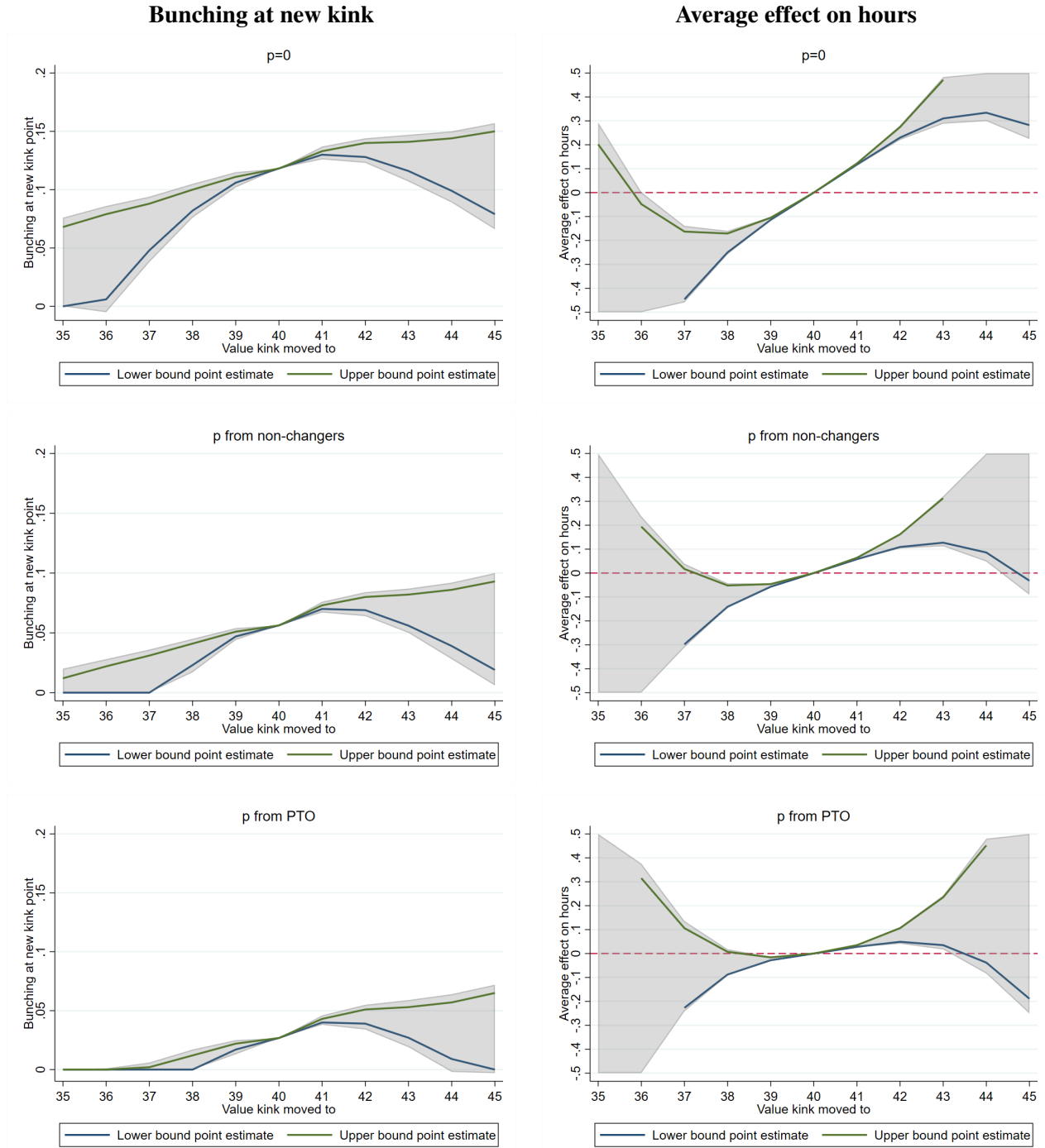
**TABLE F.7:** Hours distribution by gender, conditional on different than 40 for visibility (size of point mass at 40 can be read from Figures F.8 and F.9).

	$p=0$	$p$ from non-changers	$p$ from PTO
Net bunching:	0.090 [0.083, 0.098]	0.044 [0.041, 0.048]	0.011 [0.009, 0.012]
Buncher ATE	[1.507, 1.709] [1.387, 1.855]	[0.763, 0.814] [0.706, 0.877]	[0.187, 0.190] [0.150, 0.227]
Buncher ATE as elasticity	[0.093, 0.105] [0.086, 0.114]	[0.047, 0.050] [0.044, 0.054]	[0.012, 0.012] [0.009, 0.014]
Average effect of kink on hours	[-0.633, -0.489] [-0.688, -0.446]	[-0.319, -0.231] [-0.343, -0.213]	[-0.078, -0.054] [-0.094, -0.043]
Num observations	147953	147953	147953
Num clusters	352	352	352

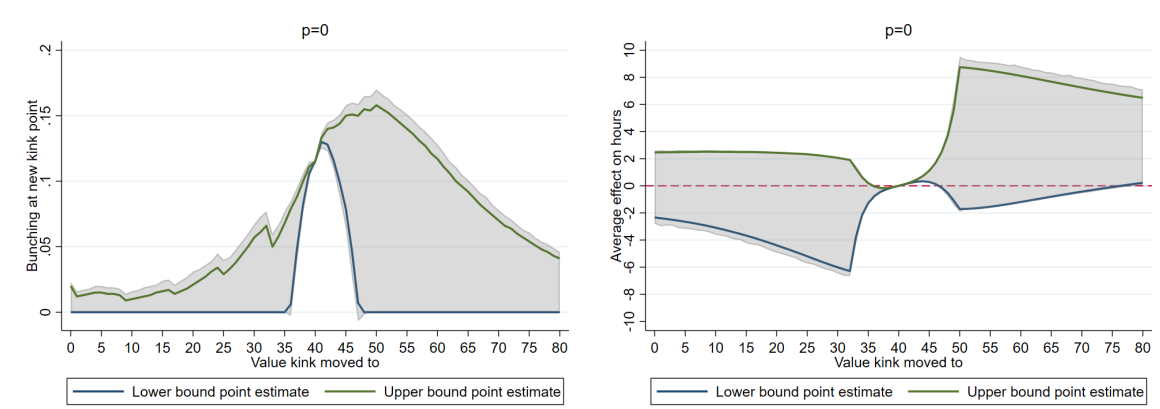
**TABLE F.8:** Hours distribution and results of the bunching estimator among women.

	$p=0$	$p$ from non-changers	$p$ from PTO
Net bunching:	0.124 [0.119, 0.129]	0.060 [0.058, 0.063]	0.031 [0.028, 0.034]
Buncher ATE	[3.074, 3.635] [2.777, 3.991]	[1.560, 1.701] [1.407, 1.869]	[0.828, 0.868] [0.717, 0.986]
Buncher ATE as elasticity	[0.190, 0.224] [0.171, 0.246]	[0.096, 0.105] [0.087, 0.115]	[0.051, 0.053] [0.044, 0.061]
Average effect of kink on hours	[-1.867, -1.271] [-2.060, -1.149]	[-0.921, -0.604] [-1.015, -0.545]	[-0.482, -0.311] [-0.549, -0.269]
Num observations	482264	482264	482264
Num clusters	524	524	524

**TABLE F.9:** Hours distribution and results of the bunching estimator among men.



**FIGURE F.9:** Bounds for the bunching that would exist at standard hours  $k$  if it were changed from 40 (left panel), as well as for the impact on average hours (right panel). Bounds of the effect on hours are clipped to the interval  $[-0.5, 0.5]$  for visibility. Pointwise bootstrapped 95% confidence intervals, cluster bootstrapped by firm, are shaded gray.



**FIGURE F.10:** Estimates of the bunching and average effect on hours were  $k$  changed to any value from 0 to 80, assuming  $p = 0$ . Bounds are not informative far from 40.

	$p=0$	$p$ from non-changers	$p$ from PTO
Net bunching:	0.116 [0.112, 0.120]	0.057 [0.055, 0.058]	0.027 [0.024, 0.030]
Treatment effect			
Linear interpolation	2.794 [2.636, 2.952]	1.360 [1.287, 1.432]	0.644 [0.568, 0.719]
Monotonicity bounds	[2.497, 3.171] [2.356, 3.353]	[1.215, 1.544] [1.153, 1.629]	[0.575, 0.731] [0.516, 0.805]
BLC buncher ATE	[2.614, 3.054] [2.493, 3.205]	[1.324, 1.435] [1.264, 1.501]	[0.640, 0.666] [0.574, 0.736]
Num observations	630217	630217	630217
Num clusters	566	566	566

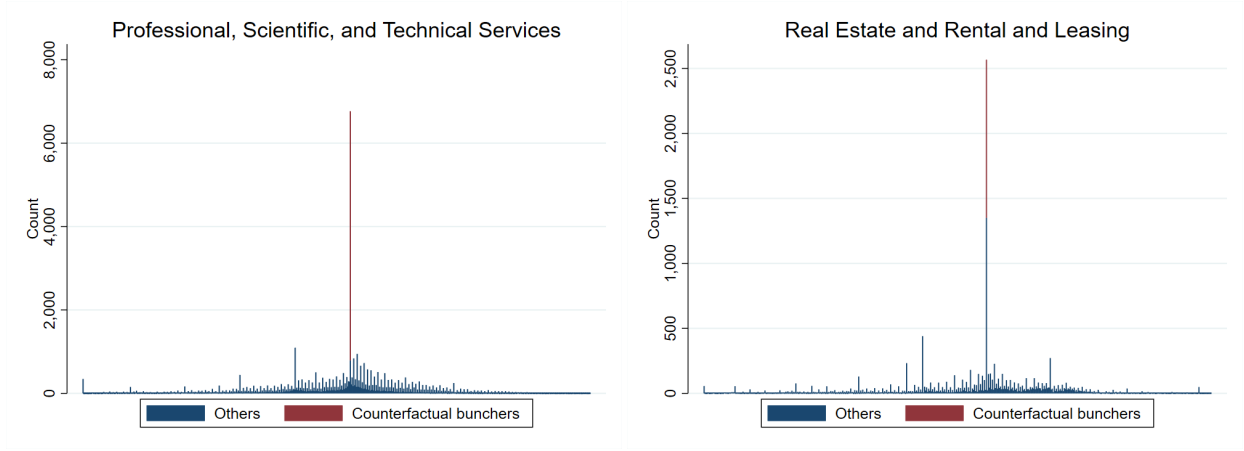
**TABLE F.10:** Treatment effects in levels with comparison to alternative shape constraints.

	$p=0$	$p$ from non-changers	$p$ from PTO
Net bunching:	0.116 [0.112, 0.120]	0.057 [0.055, 0.058]	0.027 [0.024, 0.030]
Treatment effect			
Linear interpolation	0.173 [0.163, 0.183]	0.084 [0.079, 0.088]	0.040 [0.035, 0.044]
Monotonicity bounds	[0.154, 0.196] [0.145, 0.207]	[0.075, 0.095] [0.071, 0.100]	[0.035, 0.045] [0.032, 0.050]
BLC buncher ATE	[0.161, 0.188] [0.154, 0.198]	[0.082, 0.088] [0.078, 0.093]	[0.039, 0.041] [0.035, 0.045]
Num observations	630217	630217	630217
Num clusters	566	566	566

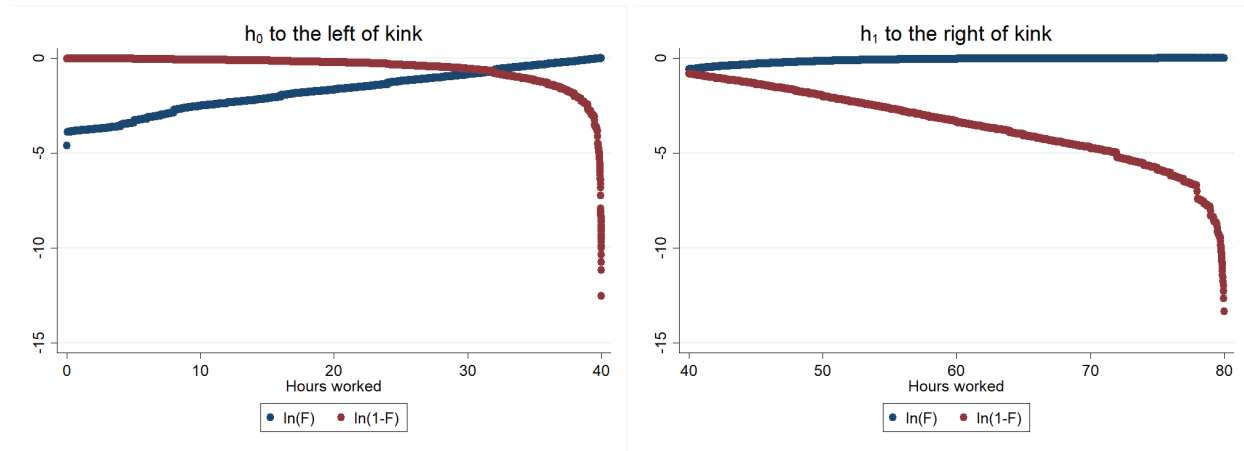
**TABLE F.11:** Treatment effects expressed as elasticities, after applying each shape constraint to the distribution of log hours rather than hours.

	$p=0$	$p$ from non-changers	$p$ from PTO
Buncher ATE as elasticity	[0.161, 0.188] [0.153, 0.198]	[0.082, 0.088] [0.077, 0.093]	[0.039, 0.041] [0.035, 0.046]
Average effect of FLSA on hours	[-1.466, -1.329] [-1.541, -1.260]	[-0.727, -0.629] [-0.769, -0.593]	[-0.347, -0.294] [-0.385, -0.262]
Avg. effect among directly affected	[-2.620, -2.375] [-2.743, -2.259]	[-1.453, -1.258] [-1.532, -1.189]	[-0.738, -0.624] [-0.814, -0.560]
Double-time, average effect on hours	[-2.604, -0.950] [-2.716, -0.904]	[-1.239, -0.492] [-1.293, -0.464]	[-0.580, -0.241] [-0.639, -0.215]

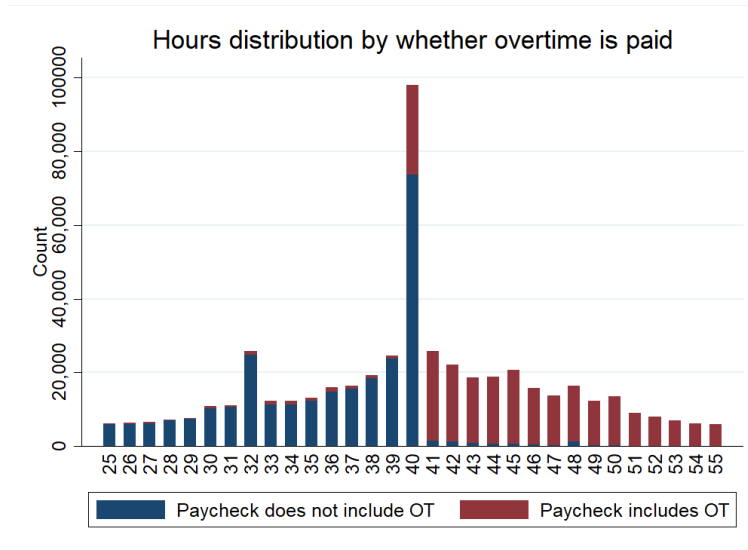
**TABLE F.12:** Estimates of policy effects (replicating Table 4) ignoring the potential effects of changes to straight-time wages.



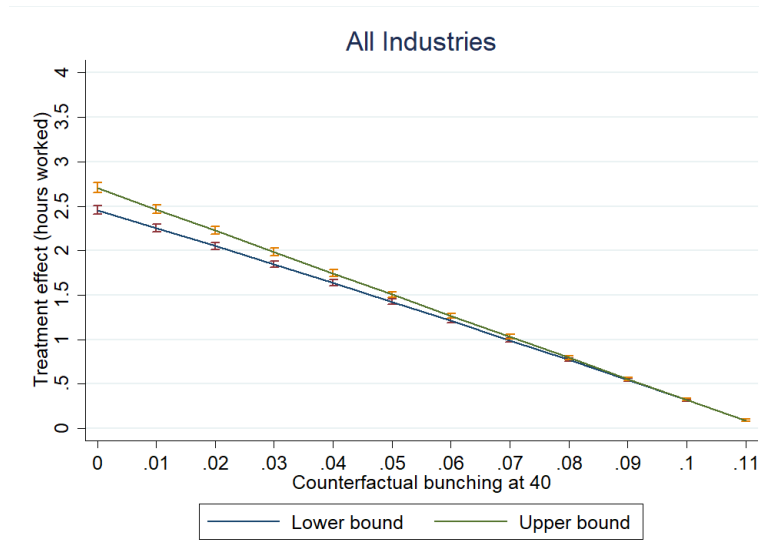
**FIGURE F.11:** Hours distribution for an industry with a low treatment effect (left), and a high one (right). Both industries exhibit a comparable amount of raw bunching (14% and 19% respectively, see Table F.6). In Professional, Scientific, and Technical Services, much more of the observable bunching is estimated to be counterfactual bunching, using the PTO-based method. Furthermore, the density of hours is higher just to the right of 40, meaning that the remaining bunching can be explained by a very small responsiveness of hours to the FLSA.



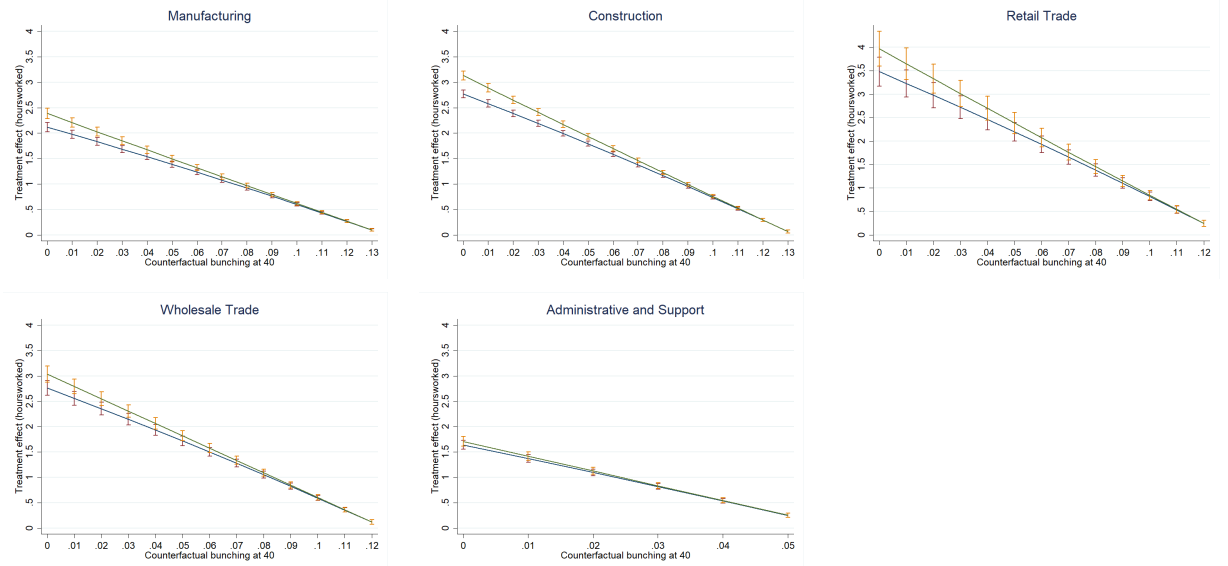
**FIGURE F.12:** Validating the assumption of bi-log-concavity away from the kink. The left panel plots estimates of  $\ln F_0(h)$  and  $\ln(1 - F_0(h))$  for  $h < 40$ , based on the empirical CDF of observed hours worked. Similarly the right panel plots estimates of  $\ln F_1(h)$  and  $\ln(1 - F_1(h))$  for  $h > k$ , where I've conditioned the sample on  $Y_i < 80$ . Bi-log-concavity requires that the four functions plotted be concave globally.



**FIGURE F.13:** Histogram of hours worked pooling all paychecks in sample, with one hour bins. Blue mass in the stacks indicate that the paycheck included no overtime pay, while red indicates that the paycheck does include overtime pay.



**FIGURE F.14:** Estimates of the buncher ATE  $\Delta_k^*$  as a function of assumed counterfactual bunching  $p$  at 40, pooling across industries. Confidence intervals depicted here are 95% intervals for each of the bounds separately.



**FIGURE F.15:** Estimates of the buncher ATE  $\Delta_k^*$  as a function of  $p$ , by each of the largest major industries.

## F.5 Estimates from the iso-elastic model

This section estimates bounds on  $\epsilon$  from the iso-elastic model under the assumption that the distribution of  $h_{0it} = \eta_{it}^{-\epsilon}$  is bi-log-concave, or linear as in Saez, 2010. If  $h_{0it}$  is BLC, bounds on  $\epsilon$  can be deduced from the fact that

$$F_0(40 \cdot 1.5^{-\epsilon}) = F_0(40) + \mathcal{B} = P(h_{it} \leq 40)$$

where  $F_0(h) := P(h_{0it} \leq h)$  and the RHS of the above is observable in the data.  $40 \cdot 1.5^{-\epsilon}$  is the location of this “marginal buncher” in the  $h_0$  distribution. In particular,

$$\epsilon = -\ln(Q_0(F_0(40) + \mathcal{B})/40)/(\ln(1.5))$$

where  $Q_0 := F_0^{-1}$  is guaranteed to exist by BLC (Dümbgen et al., 2017). In particular:

$$\epsilon \in \left[ \frac{\ln \left( 1 - \frac{1-F_0(40)}{40f(40)} \ln \left( 1 - \frac{\mathcal{B}}{1-F_0(40)} \right) \right)}{-\ln(1.5)}, \frac{\ln \left( 1 + \frac{F_0(40)}{40f(40)} \ln \left( 1 + \frac{\mathcal{B}}{F_0(40)} \right) \right)}{-\ln(1.5)} \right]$$

where  $F_0(k) = \lim_{h \uparrow 40} F(h)$  and  $f_0(k) = \lim_{h \uparrow 40} f(h)$  are identified from the data. The bounds on  $\epsilon$  estimated in this way are  $\epsilon \in [-.210, -.167]$  in the full sample.

Since BLC is preserved when the random variable is multiplied by a scalar, BLC of  $h_{0it}$  implies BLC of  $h_{1it} := \eta_{it}^{-\epsilon} \cdot 1.5^\epsilon$  as well. This implication can be checked in the data to the right of 40,



since  $\eta_{it}^{-\epsilon} \cdot 1.5^\epsilon$  is observed there. BLC of  $h_{1it}$  implies a second set of bounds on  $\epsilon$ , because:

$$F_1(40 \cdot 1.5^\epsilon) = F_1(40) - \mathcal{B} = P(h_{1it} < 40)$$

and the RHS is again observable in the data, where  $F_1(h) := P(h_{1it} \leq h)$ . Here  $40 \cdot 1.5^\epsilon$  is the location of a second “marginal buncher” – for which  $h_0 = 40$  – in the  $h_1$  distribution. Now we have:

$$\epsilon \in \left[ \frac{\ln \left( 1 + \frac{F_1(40)}{40f_1(40)} \ln \left( 1 - \frac{\mathcal{B}}{F_1(40)} \right) \right)}{\ln(1.5)}, \frac{\ln \left( 1 - \frac{1-F_1(40)}{40f_1(40)} \ln \left( 1 + \frac{\mathcal{B}}{1-F_1(40)} \right) \right)}{\ln(1.5)} \right]$$

where  $F_1(k) = F(k)$  and  $f_1(k) := \lim_{h \downarrow 40} f(h)$  are identified from the data. Empirically, these bounds are estimated as  $\epsilon \in [-.179, -.141]$ . Taking the intersection of these bounds with the range  $\epsilon \in [-.210, -.168]$  estimated previously, we have that  $\epsilon \in [-.179, -.168]$ .<sup>55</sup> The identified set is reduced from a length of .043 to .012, a factor of nearly 4. This underscores the importance of using the data from *both* sides of the kink for identification. Since a linear density satisfies BLC, the identification assumption of Saez, 2010, that the density of  $h_0$  is linear, picks a point within the identified set under BLC. Table F.10 verifies that this is born out in estimation (with results are expressed there as level effects rather than an elasticity). Table 2 reports estimates broken down by industry, as well as estimates that allow counterfactual bunching at the kink to be estimated from PTO (see Section 5).

Note that if instead of the isoelastic model we assume a more general separable and homogeneous production function like

$$\pi_{it}(z, h) = a_{it} \cdot f(h) - z$$

then treatment effects are  $\Delta_{it} = g(1/\eta_{it}) - g(1.5/\eta_{it})$ , where  $g(m) := (f')^{-1}(m)$  yields the hours  $h$  at which  $f'(h) = m$ . We can then use the fundamental theorem of calculus to express this as  $(h_{1it} - h_{0it})/h_{0it} = 1.5^{\bar{\epsilon}_{it}} - 1$  where  $\bar{\epsilon}_{it}$  is a unit-specific weighted average of the inverse elasticity of production between  $1.5\eta_{it}$  and  $\eta_{it}$ :  $\bar{\epsilon}_{it} := \int_{\eta_{it}^{-1}}^{1.5\eta_{it}^{-1}} \lambda(m) \cdot \epsilon(g(m)) \cdot dm$ , and  $\lambda(m) = \frac{1/m}{\ln 1.5}$  is a positive function integrating to one.

<sup>55</sup>Note that this interval differs slightly from the identified set of the buncher ATE as elasticity for  $p = 0$  in Table 4. The latter quantity averages the effect in levels over bunchers and rescales:  $\frac{1}{40 \ln(1.5)} \mathbb{E}[h_{0it}(1 - 1.5^\epsilon) | h_{it} = 40]$ , but the two are approximately equal under  $1.5^\epsilon \approx 1 + .5\epsilon$  and  $\ln(1.5) \approx .5$ .

## F.6 Results of the employment effect calculation

Taking my preferred estimate that FLSA eligible workers work approximately 1/3 of an hour less per week on average because of the rule, hours per worker are reduced by just under 1%. If we ignore scale effects of the overtime rule on the total number of labor hours in FLSA-eligible jobs, this suggests employment among such jobs is 1% higher than it would be without the overtime premium. This serves as an upper bound, since overall total hours worked may decrease due to overtime regulation.

Following Hamermesh (1996), assume that the percentage change in employment decomposes as:

$$\Delta \ln E|_{EH} - \eta \cdot \Delta \ln LC \cdot \frac{\eta}{\alpha - \eta} \quad (\text{F.10})$$

where  $\eta$  is constant-output demand elasticity for labor,  $\alpha$  is a labor supply elasticity. Following Hamermesh (1996) I use  $\Delta \ln LC = 0.7\%$  based on Ehrenberg and Schumann (1982), calibrated assuming that 80% of labor costs come from wages with overtime representing 2% of total hours.  $\Delta \ln E|_{EH}$  is the quantity implied by my estimates: the percentage change in employment that would occur were the total number of worker-hours  $EH$  unchanged. Taking a preferred estimate of the average effect of the FLSA as reported in Table 4 to be about 1/3 of an hour, I use a value of  $\Delta \ln E|_{EH} = \frac{1/3}{40} \approx 0.9\%$ .

		$\eta$		
		-0.15	-0.3	-0.5
$\alpha$	0	0.76	0.64	0.50
	0.1	0.80	<b>0.70</b>	0.56
	0.5	0.85	0.79	0.68

**TABLE F.13:** Back-of-the-envelope employment effects based on the average reduction in hours estimated via the bunching design and Equation (F.10), as a function of the demand elasticity for labor (rather than capital)  $\eta$ , and labor supply elasticity  $\alpha$ . The bold entry reflects “best-guess” values of  $\eta$  and  $\alpha$ .

“Best-guess” values for the other parameters used by Hamermesh, 1996 are  $\eta = -0.3$  and  $\alpha = 0.1$ , based on a review of empirical estimates. This yields 0.17 percentage points for the substitution term  $\eta \cdot \Delta \ln LC \cdot \frac{\eta}{\alpha - \eta}$ , suggesting that the effect of the FLSA is attenuated from roughly 0.87 percentage points to about a 0.70 percentage point net increase in employment—700,000 jobs assuming 100 million FLSA eligible workers.. Generating a negative overall employment response by assuming higher substitution to capital requires  $\eta = -1.25$ , well outside of empirical estimates.