Online Appendices for "Treatment Effects in Bunching Designs: The Hours Impact of the Federal Overtime Rule"

Leonard Goff

Last updated: March 5, 2022

Contents

A Identification in a generalized bunching design **2**
- A.1 The policy environment . . . 3
- A.2 Potential outcomes as counterfactual choices . . . 5
- A.3 Examples from the general choice model in the overtime setting . . . 7
- A.4 Observables in the kink bunching design . . . 9
- A.5 Treatment effects in the bunching design . . . 10
- A.6 Policy changes in the bunching-design . . . 11

B Main proofs **14**
- B.1 Proof of Lemma 1 . . . 14
- B.2 Proof of Theorem 1 . . . 16
- B.3 Proof of Theorem 2 . . . 18

C Additional empirical information and results **19**
- C.1 Sample restrictions . . . 19
- C.2 A test of the Trejo (1991) model of straight-time wage adjustment . . . 20
- C.3 Further characteristics of the sample . . . 21
- C.4 Additional treatment effect estimates and figures . . . 24
- C.5 Estimates from the iso-elastic model . . . 32
- C.6 Results of the employment effect calculation . . . 33

D Incorporating workers that set their own hours **34**

E Interdependencies among hours within the firm **38**

F Modeling the determination of wages and "typical" hours **42**
F.1 A simple model with exogenous labor supply . . . 42
F.2 Endogenizing labor supply in an equilibrium search model . . . 44
F.2.1 The model . . . 45
F.2.2 Effects of FLSA policies . . . 49
F.2.3 Calibration . . . 50

G Additional identification results for the bunching design **53**
G.1 A generalized notion of homogeneous treatment effects . . . 53
G.2 Parametric approaches with homogeneous treatment effects . . . 54
G.3 Nonparametric approaches with homogeneous treatment effects . . . 55
G.4 Alternative identification strategies with heterogeneous treatment effects . . . 56
G.5 Two bunching design settings from the literature . . . 59

H Additional proofs **61**
H.1 Proof of Propositions 1 and 2 . . . 61
H.2 Proof of Lemma 2 . . . 61
H.3 Proof of Lemma SMALL . . . 62
H.4 Proof of Appendix D Proposition 3 . . . 64
H.5 Proof of Appendix D Theorem 1* . . . 67
H.6 Proof of Appendix F.1 Proposition 4 . . . 69
H.7 Proof of Appendix Proposition 5 . . . 69
H.8 Proof of Appendix Proposition 6 . . . 70
H.9 Details of calculations for policy estimates . . . 70
H.9.1 Ex-post evaluation of time-and-a-half after 40 . . . 70
H.9.2 Moving to double time . . . 72
H.9.3 Effect of a change to the kink point on bunching . . . 73
H.9.4 Average effect of a change to the kink point on hours . . . 73
H.10 Details of calculating wage correction terms . . . 76

A Identification in a generalized bunching design

This section presents some generalizations of the bunching-design model used in the main text. 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 setting).

Further, the "running variable" of the bunching design is typically denoted by Y rather than h, and so the random variable Y_i will play the role of h_{it} from the main text. This is done to emphasize the link to the treatment effects literature, while allowing a distinction that is in some cases important (e.g. models in which hours of pay for work differ from actual hours of work).

A.1 The policy environment

Here we abstract from the conventional piece-wise linear kink setting that appears in tax examples as well as the main body of this paper. 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:

$$z \geq \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\} \tag{1}$$

The functions $B_{0i}(\mathbf{x})$ and $B_{1i}(\mathbf{x})$ are taken to be continuous and weakly convex functions of the vector $\mathbf{x}$, and assume 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.[1] In the overtime setting, $y_i(\mathbf{x})$ represents the hours of work for which a worker is paid in a given week, $k = 40$, and $B_{0i}(\mathbf{x}) = w_i y_i(\mathbf{x})$ and $B_{1i}(\mathbf{x}) = 1.5 w_i y_i(\mathbf{x}) - 20 w_i$. In most applications of the bunching design, the decision-maker $d(i)$ is simply i themself, 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 the 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 the overtime setting this means that the econometrician observes hours for which workers are paid, but not necessarily all choices made by firms that pin down those hours (for example, how many hours to allow the worker to stay "on the clock" during paid breaks–see Section A.3).

In general, 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 $B_{ki}(\mathbf{x}) := \max\{B_{0i}(\mathbf{x}), B_{1i}(\mathbf{x})\}$ denote the actual constraint function that applies to z. A budget constraint like Eq. $z \geq B_{ki}(\mathbf{x})$ is typically "kinked" because while the function $B_{ki}(\mathbf{x})$

[1] 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.

is continuous, it will generally be non-differentiable at the $\mathbf{x}$ for which $y_i(\mathbf{x}) = k$.[2] While the functions B_0, B_1 and y can all depend on i, I will often suppress this dependency for clarity of notation.

Discussion of the general model:

In the most common cases from the literature, no distinction is made between the "running variable" y of the kink and any underlying choice variables $\mathbf{x}$. This corresponds to a setting in which $\mathbf{x}$ is a scalar and $y_i(x) = x$. For example, the seminal bunching design papers Saez (2010) and Chetty et al. (2011) considered progressive taxation with z being tax liability (or credits), $y = x$ 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. Similarly, in the overtime context, the functions B_0 and B_1 are linear and only depend on hours $y_i(\mathbf{x})$, as depicted in Figure 1. Appendix G discusses a tax setting in the literature in which the functions B_0 and B_1 are linear but depend directly on a vector $\mathbf{x}$ of two components.[3] This represents a non-standard bunching-design setting, but fits naturally within the framework of this section.

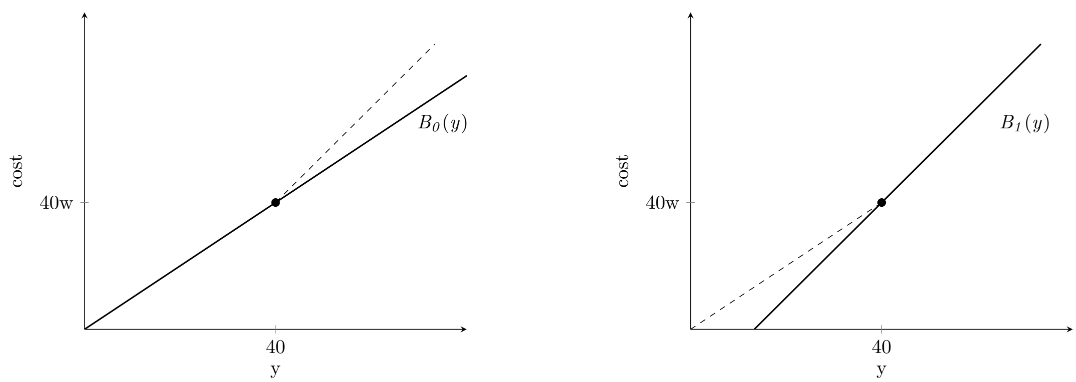


Figure 1: Definition of counterfactual cost functions B_0 and B_1 that firms could have faced, absent the overtime kink. Regardless of what choice variables are in $\mathbf{x}$, these functions only depend on $y_i(\mathbf{x})$ and are thus depicted as a function of y. Dashed lines show the rest of actual kinked-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.

Even when the functions B_0 and B_1 only depend on $\mathbf{x}$ through $y_i(\mathbf{x})$, as in standard settings, the bunching design is compatible with models in which multiple margins of choice

[2]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. With a scalar x and linear B_0 and B_1, the derivative of $B_{ki}(x)$ discontinuously rises at $\mathbf{x}$ for which $y_i(\mathbf{x}) = k$.

[3]Best et al. (2015) study firms in Pakistan that 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. In this setting a distinction between y and $\mathbf{x}$ cannot be avoided. The authors use features of the function $y_i(\mathbf{x})$ to argue that the bunching reveals changes mostly to reported costs rather than to output (see Appendix G.5 for details).

respond to the incentives provided by the kink. As discussed in the overtime context, the econometrician may be agnostic as to even what the full set of components of $\mathbf{x}$ are, with $B_{0i}(\cdot)$, $B_{1i}(\cdot)$, and $y_i(\cdot)$ depending only on various subsets of the $\mathbf{x}$ that are possibly heterogeneous by i (this is allowable because y need only be continuous in $\mathbf{x}$, and the cost functions only need to be continuous and *weakly* convex in $\mathbf{x}$, both of which are compatible with zero dependence on some of its components). Appendix G.5 gives an example in which the overtime kink gives firms an incentive to reduce bonuses, which appear in firm costs but not in the kink the variable y.

In general, the bunching design allows us to conduct causal inference on $Y_i = y_i(X_i)$, but not directly on the underlying choice variables X_i. For example in the overtime setting with possible evasion (see Sec. A.3), bunching at 40 hours will be informative about the effect of a move from B_0 to B_1 on reported hours worked y. However, it will not disentangle whether the effect on hours actually worked is attenuated by, for example, an increase in hours worked off-the-clock. The empirical setting of Best et al. (2015) provides another environment in which this point is relevant (see Appendix G.5).

A.2 Potential outcomes as counterfactual choices

Here I restate slightly more general versions of assumptions CONVEX and CHOICE from Section 4, in the present notation. As in Section 4, let us 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 $z \geq B_1(\mathbf{x})$.*

I again make explicit the assumption that these potential outcomes reflect choices made by the decision-maker. 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 $Y_{B_{0i}i}$, respectively. In this notation, the actual outcome Y_i observed by the econometrician is equal to $Y_{B_{ki}i}$.

Assumption CHOICE (perfect manipulation of y). *For any function $B(\mathbf{x})$, $Y_{Bi} = y_i(\mathbf{x}_{B_i})$, 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. Note that CHOICE here is slightly stronger than 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 except at kink, decreasing in z). *For each i and any function $B(\mathbf{x})$, choice is $(z_{Bi}, \mathbf{x}_{Bi}) = argmax_{z,\mathbf{x}}\{u_i(z, \mathbf{x}) : z \geq B(\mathbf{x})\}$ where $u_i(z, \mathbf{x})$ is weakly decreasing in z and satisfies*

$$u_i(\theta z + (1-\theta)z^*, \theta\mathbf{x} + (1-\theta)\mathbf{x}^*) > \min\{u_i(z, \mathbf{x}), u_i(z^*, \mathbf{x}^*)\}$$

for any $\theta \in (0, 1)$ and points $(z, \mathbf{x})$,$(z^*, \mathbf{x}^*)$ such that $y_i(\mathbf{x}) \neq k$ and $y_i(\mathbf{x}^*) \neq k$.

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 π as a function of the hours of a particular worker this week, and costs this week z for that worker.

Note: The second part of Assumption CONVEX is implied by strict quasi-concavity of the function $(z, \mathbf{x})$, corresponding to strictly convex preferences. However it also allows for decision-makers preferences to have "two peaks", provided that one of the peaks is located exactly at the kink. This is useful in cases in which the kink is located at a point that has particular value to decision-makers, such as firms setting weekly hours. For example, suppose that firms choose hours only $\mathbf{x} = h$, and have preferences of the form:

$$u_i(z, h) = af(h) + \phi \cdot \mathbb{1}(h = 40) - z \tag{2}$$

where $f(h)$ is strictly concave. This allows firms to have a behavioral "bias" towards 40 hours, or to extract extra profits when $h = 40$ exactly. Figure 2 depicts an example of such preferences, given an arbitrary linear budget function $B(h)$. Note that if a mass of firms were to have preferences of this form, then it would be natural to expect bunching in the distributions of h_{0it} and h_{1it}, which I allow in Section 5.

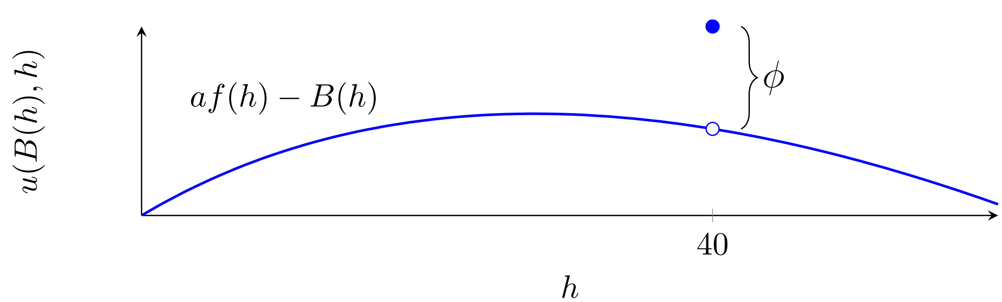


Figure 2: An example of preferences that satisfy CONVEX but are not strictly convex, cf. Eq. (2).

Note: Some departures from CONVEX are allowable without compromising it's main implication for the bunching-design, which is given in Lemma 1 below. If B_0 and B_1 are linear in $\mathbf{x}$ and the constraint $z \geq B_k(\mathbf{x})$ can be assumed to bind (hold as an equality), then the assumption that u_i is decreasing in z from CONVEX can be dropped (see Assumption CONVEX* in Appendix D). If by contrast B_0 and B_1 were strictly (rather than weakly) convex, strict convexity of preferences could be replaced with weakly convex preferences along with an assumption that u_i are strictly decreasing in z (see Eq. (5) in the Proof of Lemma 1).

Note: 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 Appendix E for details). 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.[4]

A weaker assumption than CONVEX that still has 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 (see Propositions 1 and 2 below). 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. 1 being an inequality, indicating "free-disposal").

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

To demonstrate the flexibility of the general choice model CONVEX, I below present some examples for the overtime setting. These examples are meant only to be illustrative, and each could apply to a different subset of units in the population. In these examples we continue

[4] I note that SUTVA issues like those addressed in Appendix E 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.

to take the decision-maker for a given unit to be the firm employing that worker.[5]

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 $h_m := \text{argmax}_h f(h) - \nu(h)$, provided that the corresponding optimal bonus is non-negative. Bonuses fully adjust to counteract overtime costs, and $h_0 = h_1 = h_m$.

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 $\pi(z, h, o) = f(h) - \phi(o) - z$.[6] 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 isoelastic in a CES aggregate (g could also denote planned hours next week): $\pi(z, h, g) = a \cdot \left((\gamma h^\rho + g^\rho)^{1/\rho}\right)^{1+\frac{1}{\epsilon}} - z$ with γ 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 $f(h) = a \cdot \left((\gamma h^\rho + g^{*\rho})^{1/\rho}\right)^{1+\frac{1}{\epsilon}}$. 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 ϵ.[7] Section 4.4 discusses how complementaries affect the final evaluation of the FLSA.

[5] Appendix D 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.

[6] 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.

[7] 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 $\rho < 1 + 1/\epsilon$.

A.4 Observables in the kink bunching design

Lemma 1 outlines the core consequence of Assumption CONVEX 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:*

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

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 be seen as an observation of one or the other potential outcome, or k, depending on how the potential outcomes relate to the kink point k.

Note that the "straddling" event $Y_{0i} \leq k \leq Y_{1i}$ from Lemma 1 can be written as $Y_{0i} \in [k, k + \Delta_i]$, where $\Delta_i := Y_{0i} - Y_{1i}$. Similarly, we can 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*.

Let $\mathcal{B} := P(Y_i = k)$ be the observable probability that the decision-maker chooses to locate exactly at $Y = k$. Proposition 1 gives the relationship between this bunching probability and treatment effects, which holds in a weakened form when CONVEX is replaced by WARP:

Proposition 1 (relation between bunching and Δ_i). *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 H. □

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.[8] Let $F_1(y) = P(Y_{0i} \leq y)$ 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.*

[8] Note that in the overtime application sampling is actually at the firm level, which coincides with the level of decision-making units $d(i)$.

Proof. See Appendix H. □

As an example of how WARP alone (without CONVEX) can still be useful for identification, suppose that $\Delta_i = \Delta$ were known to be homogenous across units,[9] and $f_0(y)$ were constant across the interval $[k, k+\Delta]$, then by Propositions 1 and 2 we have that $\Delta \geq \mathcal{B}/f_0(k)$ under WARP and CHOICE.

A.5 Treatment effects in the bunching design

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

Unit i's treatment effect $\Delta_i := Y_{0i} - Y_{1i}$ 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. These treatment effects 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 Δ_i (Appendix G.5 discusses this in the setting of Best et al. 2015). 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 revealed by bunching at a kink, complicating efforts to identify a elasticity parameter having a firm structural interpretation (Einav et al., 2017).

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 the note of caution by Einav et al. (2017), which points out 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 only as an intermediate step in the logic, which only requires the nonparametric restrictions of convexity and monotonicity rather than knowing its functional form (or even what vector of choice variables $\mathbf{x}$ underly a given agent's observed value of y). This greatly increases the robustness of the method to potential misspecification of the underlying choice model.

[9] One way to get homogenous treatment effects in levels in the overtime setting is to assume exponential production: $f(h) = \gamma(1-e^{-h/\gamma})$ where $\gamma > 0$ and $h_{0it} - h_{1it} = \gamma \ln(1.5)$ for all units. The iso-elastic model instead gives homogeneous treatment effects for $log(h)$.

Additional identification results for the bunching design:

While Theorem 1 of Section 4 developes the treatment effect identification result used to evaluate the FLSA, Appendix G 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 canonical results from the literature 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). Appendix G also discusses 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, but continues with the notation of this Appendix. When RANK fails (and $p = 0$ for simplicity), the bounds from Theorem 1 are still valid under BLC of Y_0 and Y_1 for the following averaged quantile treatment effect:

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

where $\Delta_0^* := Q_0(F_1(k)) - Q_1(F_1(k)) = Q_0(F_1(k)) - k$ and $\Delta_1^* := Q_0(F_0(k)) - Q_1(F_0(k)) = k - Q_1(F_0(k))$. Thus, Δ_0^* is the value such that $F_0(k + \Delta_0^*) = F_0(k) + \mathcal{B}$, and Δ_1^* is the value such that $F_1(k - \Delta_1^*) = F_1(k) - \mathcal{B}$. The averaged quantile treatment effect of Eq. (3) yields a lower bound on the buncher ATE, as described in Fig. 3.

A.6 Policy changes in the bunching-design

This section presents the logic establishing Theorem 2 in the main text regarding the effects of changes to the policy generating a kink. Consider a bunching design setting 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 ρ 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 ρ represents a wage-scaling factor, with $\rho = 1$ for straight-time and $\rho = 1.5$ for overtime:

$$B_i(y; \rho, k) = \rho w_i y - k w_i(\rho - 1) \quad (4)$$

where work hours y may continue to be a function $y(\mathbf{x})$ of a vector of choice variables to the firm. In this example, k controls the size of the lump-sum subsidy $k w_i(\rho - 1)$ that keeps $B_i(k; \rho, k)$ invariant as ρ is changed.

Signing the bias when RANK fails

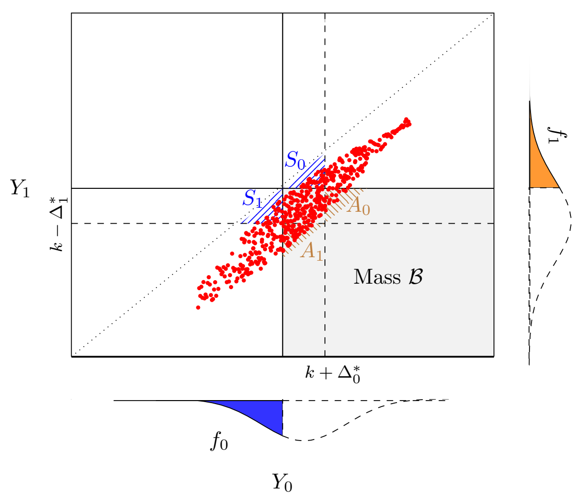


Figure 3: When Assumption RANK fails, the average $\mathbb{E}[Y_{0i}|Y_{0i} \in [k, k + \Delta_0^*]]$ will include the mass in the region S_0, who are not bunchers (NE lines) but will be missing the mass in the region A_0 (NW lines) who are. This causes an under-estimate of the desired quantity $\mathbb{E}[Y_{0i}|Y_{1i} \le k \le Y_{0i}]$. Similarly, $\mathbb{E}[Y_{1i}|Y_{1i} \in [k - \Delta_1^*, k]]$ 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 $\mathbb{E}[Y_{1i}|Y_{1i} \le k \le Y_{0i}]$.

In the general setting, assume that ρ takes values in a convex subset of $\mathbb{R}$ containing ρ_0 and ρ_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 ρ and k). 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 (4) satisfies these requirements with $y_i(h) = h$.[10]

For any value of ρ, 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 $z \ge B_i(\mathbf{x}; \rho, k)$. 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 the overtime setting with Equation (4), along with quasi-linearity of preferences. Quasilinearity of preferences is a property of profit-maximizing firms when z represents a cost, and is thus a natural assumption in the overtime setting.

Assumption SEPARABLE (invariance of potential outcomes with respect to k). *For all i, ρ 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 ϕ_i), and for all i $u_i(z, \mathbf{x})$ can be chosen to be additively separable and linear in z.*

[10] As an alternative example, I construct in Appendix G.5 functions $B_i(\mathbf{x}; \rho, k)$ for the bunching design setting from Best et al. (2015). In that case, ρ 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 ρ.

Additive separability of $B_i(\mathbf{x};\rho,k)$ in k may be context specific: in the example from Best et al. (2015) described in Appendix G.5, 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 ρ_1 or k on hours. Lemma 2 generalizes an existing result from Blomquist et al. (2015), and makes use of a regularity condition I introduce in the proof as Assumption SMOOTH.[11] Counterfactual bunchers $K_i^* = 1$ are assumed to stay at some fixed value k^* (40 in the overtime setting), regardless of ρ 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 expressed in terms of marginal responsiveness). *Assume CHOICE, SMOOTH and WARP. Then:*

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

with equality under CONVEX.

Proof. See Appendix H. □

The main tool in establishing Lemma 2 is to 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.

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

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

Proof. See Appendix H. □

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 $B_i(\mathbf{x};\rho',k)$ (under WARP, with equality under CONVEX). As a result, Lemma SMALL shows that the uniform density approximation that has appeared in Saez (2010) and Kleven (2016) (stated in Appendix Theorem 8) for "small" kinks becomes exact in the limit that the two cost functions approach one another. The small kink approximation says that $\mathcal{B} \approx f_\rho(k) \cdot \mathbb{E}[Y_i(\rho) - Y_i(\rho')]$, where note that treatement effects can be writtens:

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

[11] Blomquist et al. (2021) derive the special case of Lemma 2 with convex preferences over a scalar choice variable and $p = 0$, in the context of labor supply under piecewise linear 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.

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 ρ_1 is increased, under CONVEX.

Lemma 2 is useful for identification results regarding changes to k when it is combined with a result from Kasy (2022), 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 (continuous flows under a small change to ρ). *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 Kasy (2022). □

The intuition behind Lemma 3 comes from the physical dynamics of fluids. When ρ 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 ρ will lead to a change in the density there.

With ρ_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 (see Appendix B for proof), 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 $Y_{1i} \le k \le Y_{0i}$. However, to finish establishing Lemma 1, we also need the reverse implication: that $Y_{1i} \le k \le Y_{0i}$ implies $Y_i = k$. This comes from taking the contrapositive of each of the two claims in item ii).

Proof of i): Let $\mathcal{X}_{0i} = \{\mathbf{x} : y_i(\mathbf{x}) \le k\}$ and $\mathcal{X}_{1i} = \{\mathbf{x} : y_i(\mathbf{x}) \ge k\}$. If $Y_{0i} \le k$, then by CHOICE $\mathbf{x}_{B_0}$ is in $\mathcal{X}_0$. Since $B_k(\mathbf{x}) = B_0(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}_0$, it follows that $z_{B_0 i} \ge B_k(\mathbf{x}_{B_0 i})$, i.e. Y_{0i} is feasible under B_k. Note that $B_{ki}(\mathbf{x}) \ge B_{0i}(\mathbf{x})$ for all $\mathbf{x}$. By WARP then $(z_{B_k i}, \mathbf{x}_{B_k i}) = (z_{B_0 i}, \mathbf{x}_{B_0 i})$. Thus $Y_i = y_i(\mathbf{x}_{B_k}) = y_i(\mathbf{x}_{B_0}) = Y_{0i}$. So $Y_{0i} \le k \implies Y_i = Y_{0i}$. By the same logic we can show that $Y_{1i} \ge k \implies Y_i = Y_{1i}$.

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

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

Suppose that $y_i(\mathbf{x}_{Bi}) \ne k$, and consider any $\mathbf{x} \ne \mathbf{x}_{Bi}$ such that $y_i(\mathbf{x}) \ne k$. 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:

$$u_i(B(\tilde{\mathbf{x}}), \tilde{\mathbf{x}}) \ge u_i(\theta B(\mathbf{x}) + (1-\theta)B(\mathbf{x}^*), \tilde{\mathbf{x}}) > \min\{u_i(B(\mathbf{x}), \mathbf{x}), u_i(B(\mathbf{x}^*), \mathbf{x}^*)\} = u_i(B(\mathbf{x}), \mathbf{x}) \tag{5}$$

where I have used CONVEX in the second step, and that $\mathbf{x}^*$ is a maximizer in the third. This result implies that for any such $\mathbf{x} \ne \mathbf{x}^*$, 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 A1 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 1* in which some workers choose their hours.

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

$$u_{B_k i}(\mathbf{x}) = \begin{cases} u_{B_0 i}(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{X}_{0i} \\ u_{B_1 i}(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{X}_{1i} \end{cases}$$

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 $Y_i < k$. Suppose furthermore that $Y_{0i} \ne Y_i$, with $Y_{0i} = y_i(\mathbf{x}_{0i})$ and $\mathbf{x}_{0i}$ the maximizer of $u_{B_0 i}(\mathbf{x})$. Note that we must have that $\mathbf{x}_{0i} \notin \mathcal{X}_{0i}$, because $B_0 = B_k$ in $\mathcal{X}_{0i}$ so we can't have $u_{B_0 i}(\mathbf{x}_{0i}) > u_{B_0 i}(\mathbf{x}_{ki})$ (since $\mathbf{x}_{ki}$ maximizes $u_{B_k i}(\mathbf{x})$). Thus $Y_{0i} > k$.

By continuity of $y_i(\mathbf{x})$, $\mathcal{X}_{0i}$ is a closed set and $\mathbf{x}_{ki}$ belongs to the interior of $\mathcal{X}_{0i}$. Thus, while $\mathbf{x}_{0i}$ is not in $\mathcal{X}_{0i}$, there exists a point $\tilde{\mathbf{x}} \in \mathcal{X}_{0i}$ along the line between $\mathbf{x}_{0i}$ to $\mathbf{x}_{ki}$. Since $Y_i \neq k$ and $Y_{0i} \neq k$, Eq. (5) with $B = B_k$ then implies that $u_{B_k i}(\tilde{\mathbf{x}}) > u_{B_k i}(\mathbf{x}_{0i})$. Since $u_{B_0 i}(\mathbf{x}) = u_{B_k i}(\mathbf{x})$ for all $\mathbf{x}$ in $\mathcal{X}_{0i}$, it follows that $u_{B_0 i}(\tilde{\mathbf{x}}) > u_{B_0 i}(\mathbf{x}_{0i})$. However, this contradicts the premise that $\mathbf{x}_{0i}$ maximizes $u_{B_0 i}(\mathbf{x})$. Thus, $Y_i < k$ implies $Y_i = Y_{0i}$. Figure 4 depicts the logic visually. The proof that $Y_i > k$ implies $Y_i = Y_{1i}$ is analogous.

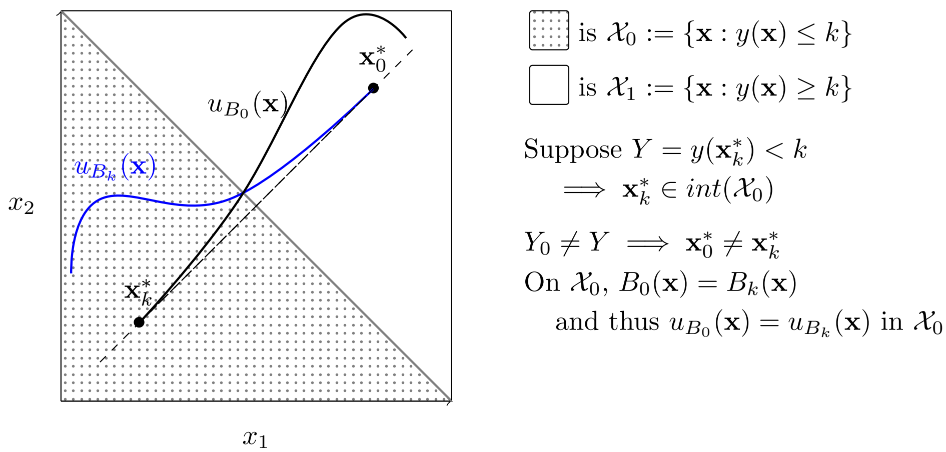


Figure 4: Depiction of the step establishing $(Y < k) \implies (Y = Y_0)$ in the proof of Lemma 1. In this example $z = (x_1, x_2)$ and $y(\mathbf{x}) = x_1 + x_2$. We suppress indices i for clarity. Proof is by contradiction. If $Y_0 \neq Y$, then $\mathbf{x}_k^* \neq \mathbf{x}_0^*$, where $\mathbf{x}_k^*$ and $\mathbf{x}_0^*$ are the unique maximizers of $u_{B_k}(\mathbf{x})$ and $u_{B_0}(\mathbf{x})$, respectively. By Equation 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 $u_{B_0}(\mathbf{x})$, 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 z_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 $\mathcal{X}_0$. 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.

B.2 Proof of Theorem 1

Theorem 1 of Dümbgen 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, given Assumption RANK and Lemma 1:

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

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

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

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:

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

where

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

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:

$$F_{0|K^*=0}(k) = \frac{F_0(k)-p}{1-p} \quad \text{and} \quad F_{1|K^*=0}(k) = \frac{F_1(k)-p}{1-p} \quad \text{and} \quad \mathcal{B}^* = \frac{\mathcal{B}-p}{1-p}$$

$$f_{0|K^*=0}(k) = \frac{f_0(k)}{1-p} \quad \text{and} \quad f_{1|K^*=0}(k) = \frac{f_1(k)}{1-p}$$

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 Proposition 2 and the BLC assumption, the above quantities are related to observables as:

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

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 the theorem.

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

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

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

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.3 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 Appendix A Lemmas 2 and 3 the effect of changing k on bunching is:

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

Turning now to the total effect on average hours.

$$\begin{aligned}\partial_k E[Y_i^{[k,\rho_1]}] &= \partial_k\left\{P(Y_i(\rho_0)<k)\mathbb{E}[Y_i(\rho_0)|Y_i(\rho_0)<k]\right\} + k\partial_k\left(\mathcal{B}^{[k,\rho_1]}-p(k)\right) + \mathcal{B}^{[k,\rho_1]}-p(k) \\ &\qquad + \partial_k\left\{P(Y_i(\rho_1)>k)\mathbb{E}[Y_i(\rho_1)|Y_i(\rho_1)>k]\right\} \\ &= \partial_k\int_{-\infty}^{k} y\cdot f_{\rho_0}(y)\cdot dy + k\left(f_0(k)-f_1(k)\right) + \mathcal{B}^{[k,\rho_1]} - p(k) + \partial_k\int_{k}^{\infty} y\cdot f_{\rho_1}(y)\cdot dy \\ &= \cancel{kf_0(k)} + \cancel{k\left(f_1(k)-f_0(k)\right)} + \mathcal{B}^{[k,\rho_1]} - p(k) - \cancel{kf_1(k)}\end{aligned}$$

Meanwhile: $\partial_{\rho_1}\mathbb{E}[Y_i^{[k,\rho_1]}] = -\int_k^\infty f_{\rho_1}(y)\mathbb{E}\left[\frac{dY_i(\rho_1)}{d\rho}\middle| Y_i(\rho_1) = y\right] dy$ follows directly from Lemma 2 and differentiating both sides with respect to ρ_1, and thus

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

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 $\lim_{y\to\infty} y \cdot f_{\rho_1}(y) = 0$, implied by Assumption SMOOTH.

C Additional empirical information and results

C.1 Sample restrictions

Beginning with the initial sample described in Column (2) of Table 1, 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.[12] 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, the assumption being 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 that they appear in the data,[13] 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.

[12]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.

[13]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.

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

One 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). Given the kink in Eq. (1), we can perform such a test using the wage and hours reported on each paycheck, while making only weak differentiability assumptions on unobservables for identification.

Suppose that for some subset of units it, wages are actively adjusted to the hours they work according to Equation (1), in order to target some total earnings z_{it}. Denote the corresponding units by a latent variable $A_{it} = 1$. These units may come from workers with limited variation in their schedules in those weeks in which $h_{it} = h_i^*$ for some typical hours h_i^* according to which their wages were set by Eq. (1) at hiring. $A_{it} = 1$ units might instead have dynamic wages that adjust to week-by-week variation in their hours h_{it}. 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:

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

The middle term above introduces a kink in the conditional expectation of log wages with respect to hours. If we assume that $\mathbb{E}\left[\ln z_{it}|h_{it} = h, A_{it} = 1\right]$, $\mathbb{E}\left[\ln w_{it}|h_{it} = h, A_{it} = 0\right]$ 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 $h = 40$:

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

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 at 40. 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 5 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 $q(40) \approx 0.25$ 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.

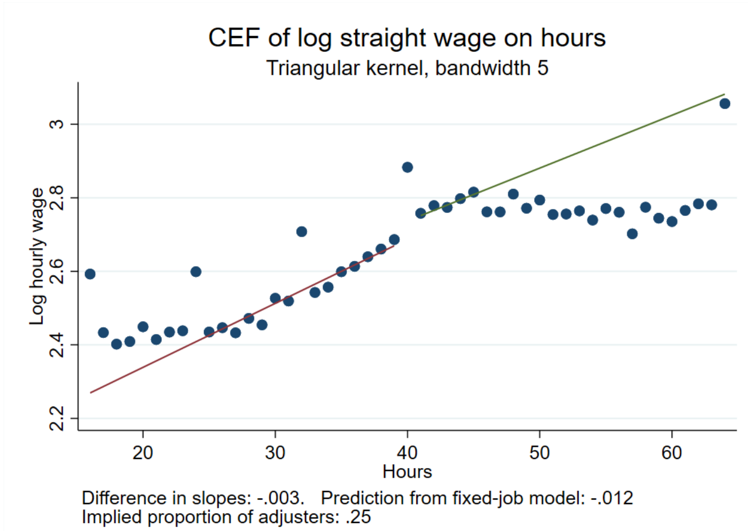


Figure 5: A kinked-CEF test of the fixed-jobs model presented in Trejo (1991). Regression lines fit on each side with a uniform kernel within 25 hours of the 40.

C.3 Further characteristics of the sample

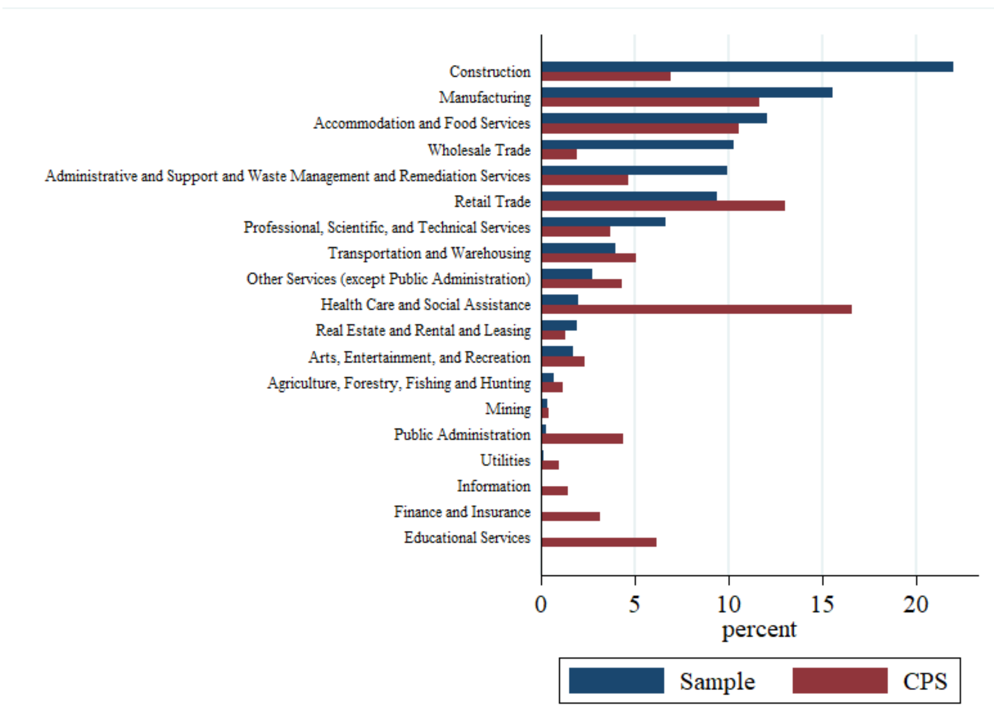


Figure 6: Industry distribution of estimation sample versus the Current Population Survey sample described in Section 3.

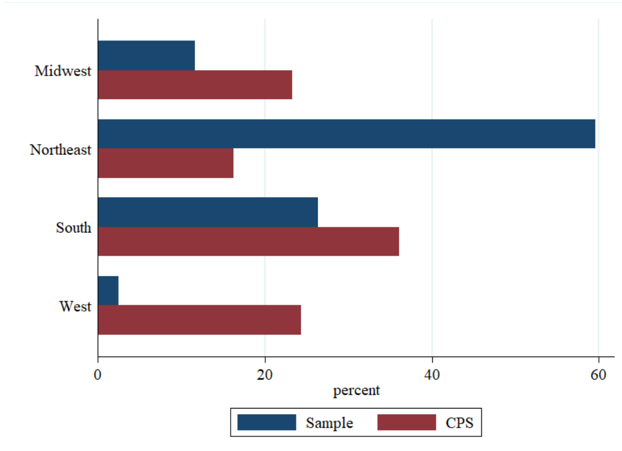


Figure 7: Geographical distribution of estimation sample versus the Current Population Survey sample described in Section 3.

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 1: 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	0.0515	0.0796		
	(0.95)	(3.95)	(3.31)		
Age	0.000690	0.00266	0.0250		
	(3.82)	(0.74)	(3.25)		
Female	0.0140	-1.322	-1.943		
	(2.08)	(-9.07)	(-6.08)		
Minimum wage worker	0.00121	-1.687	-5.352		
	(0.29)	(-2.39)	(-4.08)		
Firm just hired				-0.00572	0.553
				(-2.95)	(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 2: 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

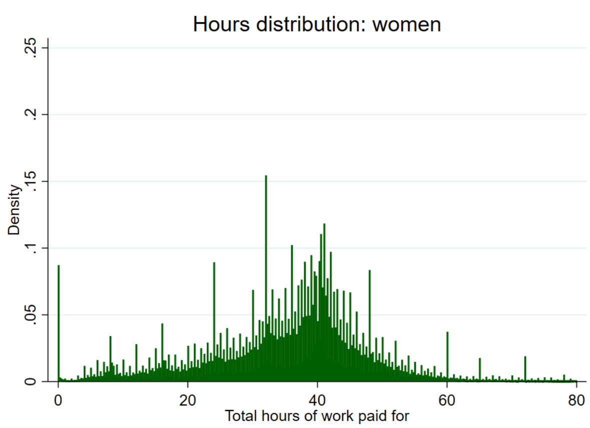
t statistics in parentheses

Table 3: Decomposing variation in total hours. Worker fixed effects and employer by day fixed effects explain about 63% of the variation in total hours.

C.4 Additional treatment effect estimates and figures

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

Table 4: 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.



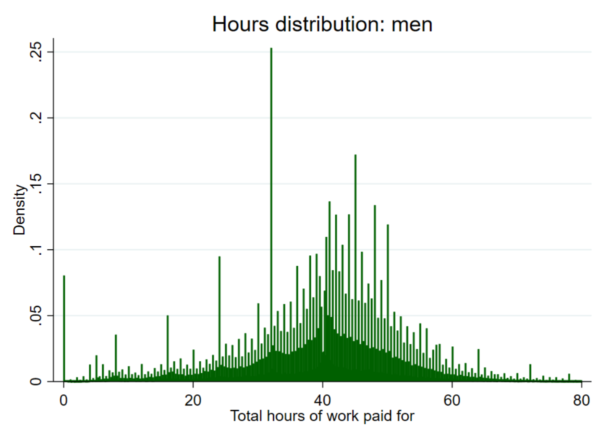


Table 6: Hours distribution by gender, conditional on different than 40 for visibility (size of point mass at 40 can be read from Figures 7 and 8).

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

Table 5: 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 $p = 0$.

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

Table 7: Results of the bunching estimator among women.

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

Table 8: Results of the bunching estimator among men.

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

Table 9: Treatment effects in levels with comparison to alternative shape constraints.

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

Table 10: Treatment effects expressed as elasticities, after applying each shape constraint to the distribution of log hours rather than the distribution of hours.

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

Table 11: Estimates of policy effects (replicating Table 4) ignoring the potential effects of changes to straight-time wages.

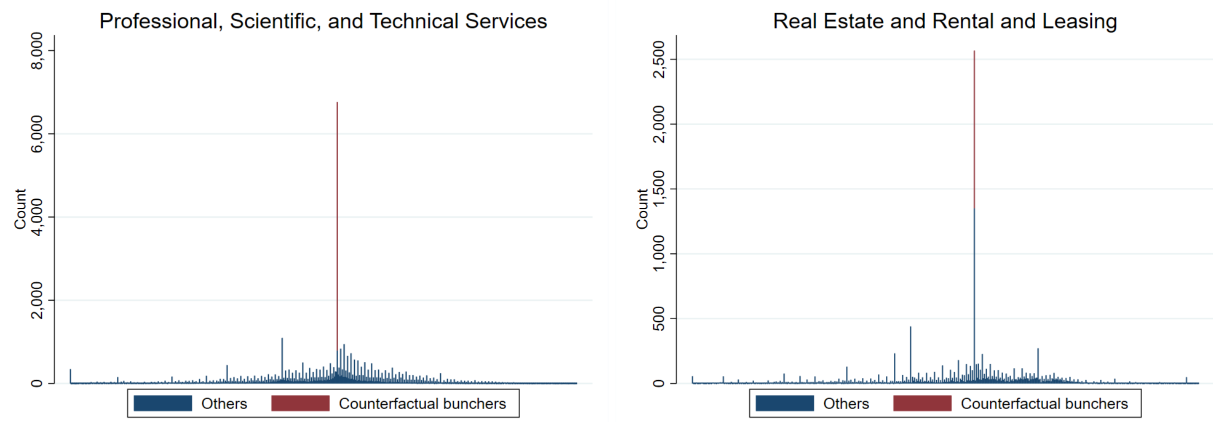


Figure 8: Hours distribution for an industry with a low treatment effect (left), and a high one (right). Both industries exhibit a comparable amount of raw bunching (14% and 19% respectively, see Table 5). In Professional, Scientific, and Technical Services, much more of the observable bunching is estimated 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.

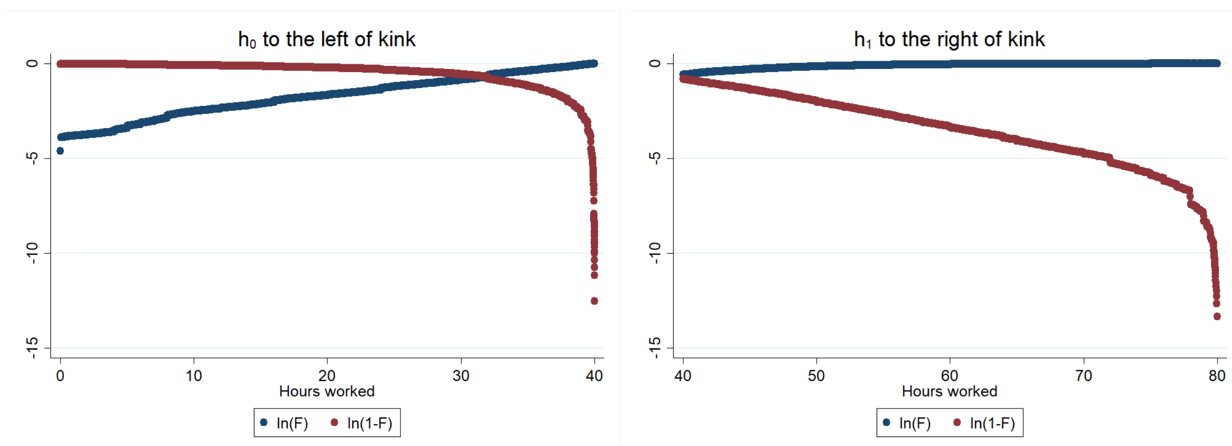


Figure 9: Validating the assumption of bi-log-concavity away from the kink. The left panel plots estimates of $\ln F_0(h)$ and $\ln(1 - F_0(h))$ for $h < 40$, based on the empirical CDF of observed hours worked. Similarly the right panel plots estimates of $\ln F_1(h)$ and $\ln(1 - F_1(h))$ for $h > k$, where I've conditioned the sample on $Y_i < 80$. Bi-log-concavity requires that the four functions plotted be concave globally.

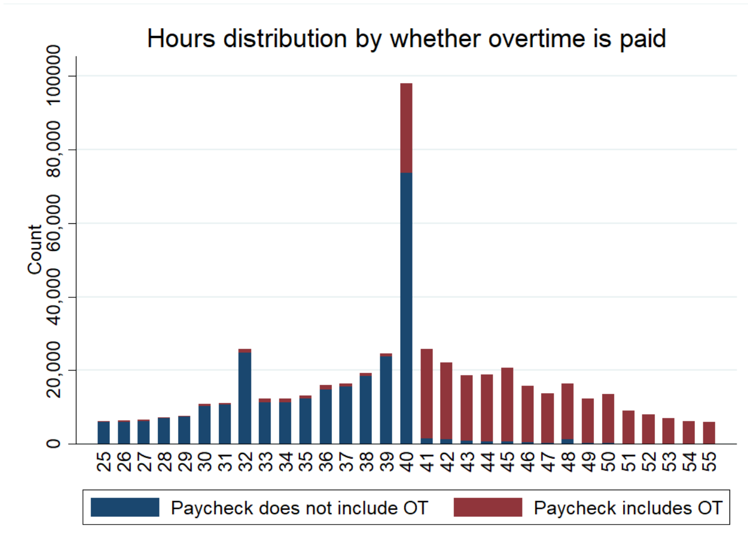


Figure 10: Histogram of hours worked pooling all paychecks in sample, with one hour bins. Blue mass in the stacks indicate that the paycheck included no overtime pay, while red indicates that the paycheck does include overtime pay.

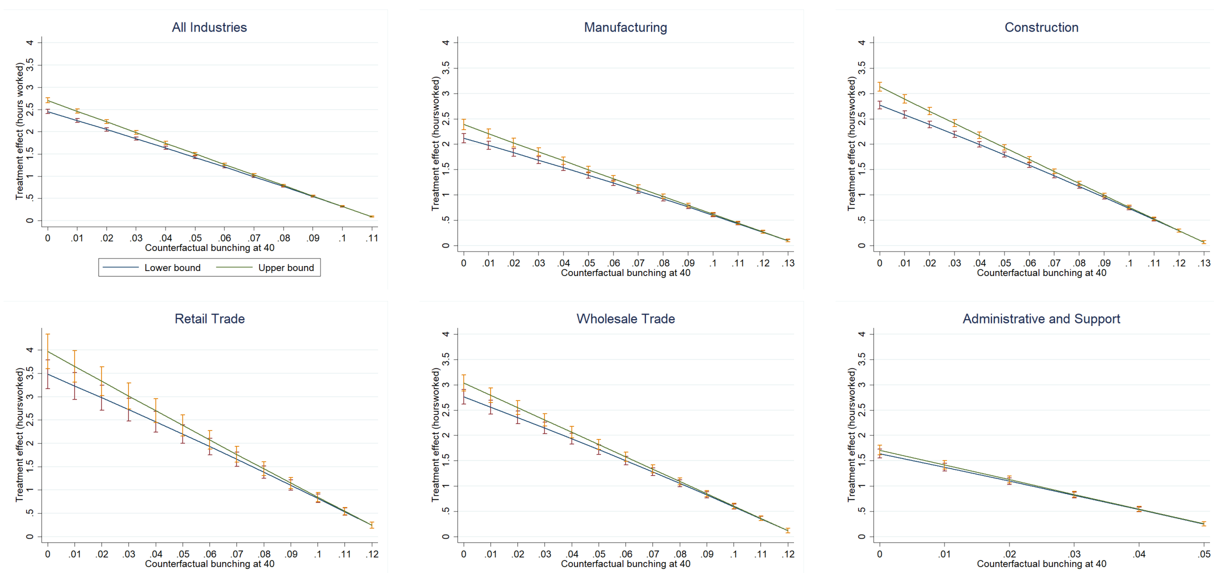


Figure 11: Estimates of the buncher ATE Δ_k^* as a function of p, pooled across industries and by each of the largest major industries.

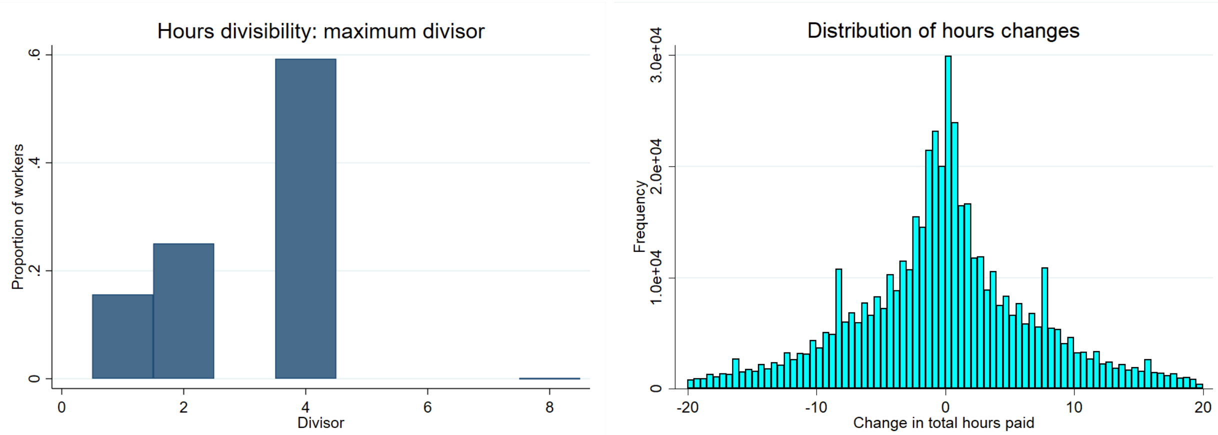


Figure 12: *Left:* distribution of the largest integer $m = 1 \ldots 10$ that maximizes the proportion of worker i's paychecks for which hours are divisible by m. This can be thought of as the granularity of hours reporting for worker i. *Right:* distribution of changes in total hours between subsequent pay periods (truncated at -20 and 20)

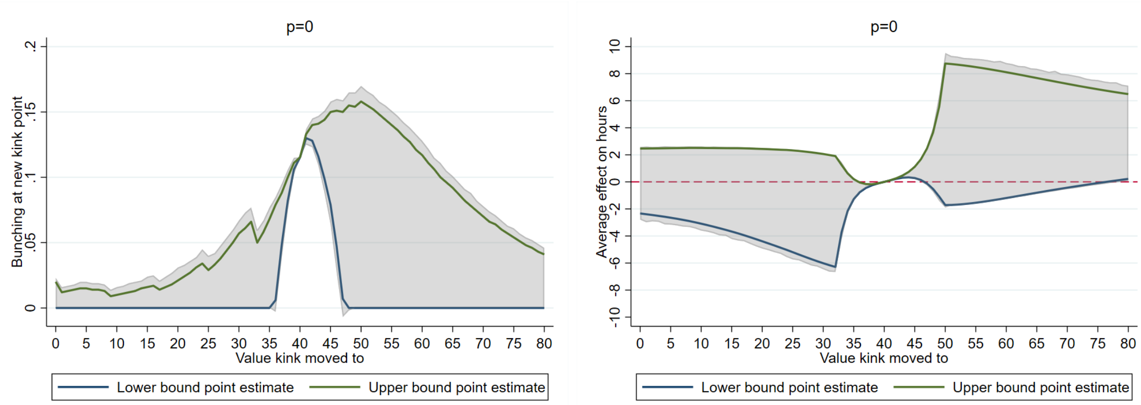


Figure 13: Estimates of the bunching (left panel) and average effect on hours (right panel) were k changed to any value from 0 to 80, assuming $p = 0$. Pointwise bootstrapped 95% confidence intervals, cluster bootstrapped by firm, are shaded gray. Bounds are not informative far from 40.

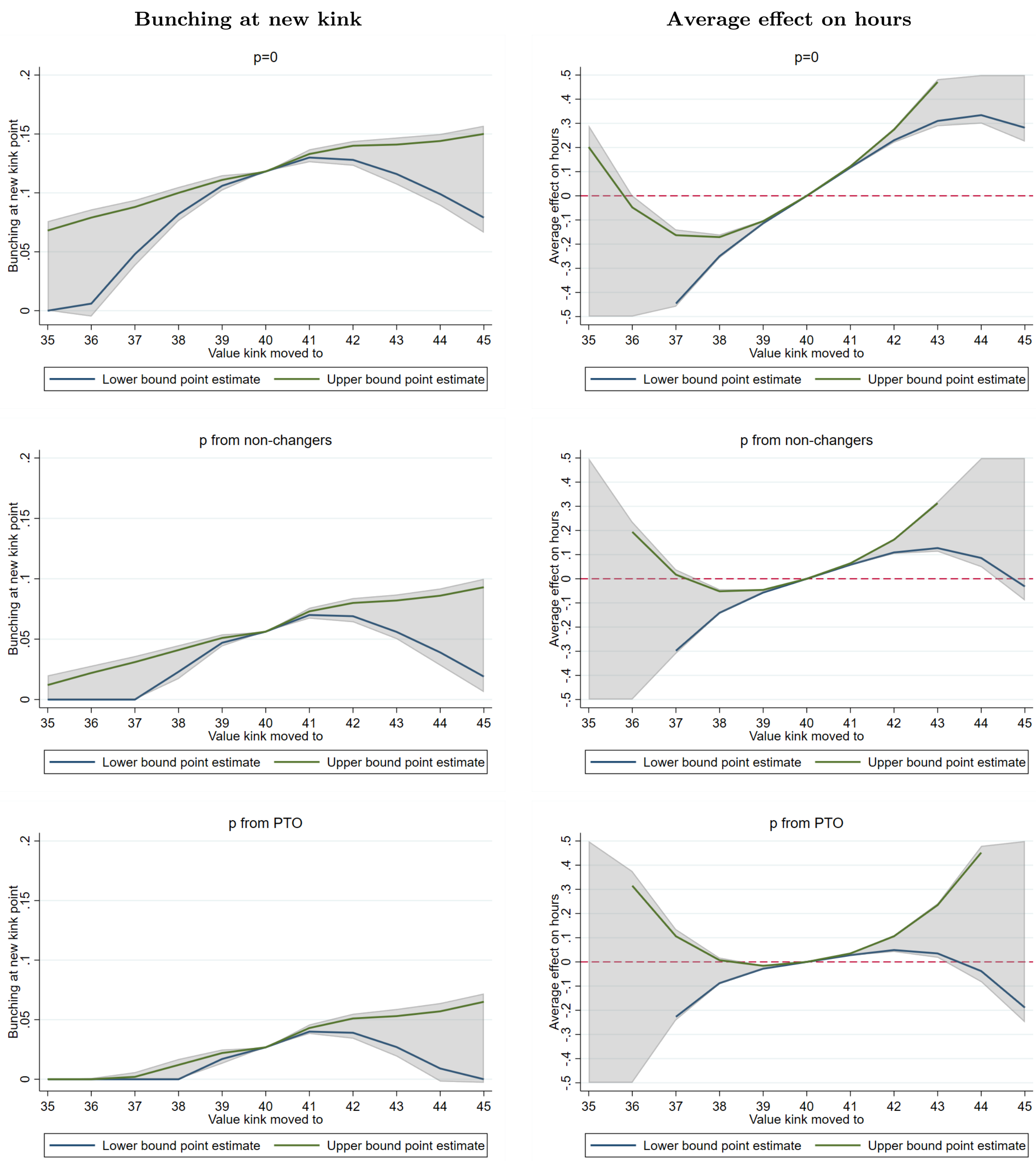


Figure 14: Bounds for the bunching that would exist at standard hours k if it were changed from 40 (left panel), as well as for the impact on average hours (right panel). Bounds of the effect on hours are clipped to the interval $[-0.5, 0.5]$ for visibility. Pointwise bootstrapped 95% confidence intervals, cluster bootstrapped by firm, are shaded gray.

C.5 Estimates from the iso-elastic model

This section estimates bounds on ϵ from the iso-elastic model described in Section 4.2, under the assumption that the distribution of $h_{0it} = \eta_{it}^{-\epsilon}$ is bi-log-concave (and linear as in Saez, 2010 as a special case). If h_{0it} is BLC, bounds on ϵ 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 ϵ estimated in this way are $\epsilon \in [-.210, -.167]$ in the full sample, with all bunching $\mathcal{B}$ attributed to the kink ($p = 0$).

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 ϵ, because:

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

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 $\epsilon \in [-.179, -.168]$.[14] 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

[14]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 $\ln(1.5) \approx .5$.

linear, picks a point within the identified set under BLC. Table 9 verifies that this is born out in estimation (with results are expressed there as level effects rather than an elasticity).

As discussed in Section 4.2, a value fo $\epsilon \approx -.175$ is difficult to reconcile with a realistic view of revenue production with respect to hours. Note that if instead of the isoelastic model, production were instead described by 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 η_{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. Here $\bar{\epsilon}_{it}$ plays the role of an "effective" elasticity parameter that determines the size of treatment effects when the production function is $f(h)$. This suggests that simply generalizing the functional form $f(h)$ is not sufficient to reconcile a realistic picture of production with the data, since the observed bunching still maps to a local average elasticity of $f(h)$. However, the general choice model that allows multiple margins of choice $\mathbf{x}$ can.

C.6 Results of the employment effect calculation

Taking my prefered 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:

$$\Delta \ln E|_{EH} - \eta \cdot \Delta \ln LC \cdot \frac{\eta}{\alpha - \eta} \tag{6}$$

where η is constant-output demand elasticity for labor, α 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 $\Delta \ln E|_{EH} = \frac{1/3}{40} \approx 0.9\%$.

"Best-guess" values for the other parameters used by Hamermesh, 1996 are $\eta = -0.3$ and

		η		
		-0.15	-0.3	-0.5
	0	0.76	0.64	0.50
α	0.1	0.80	**0.70**	0.56
	0.5	0.85	0.79	0.68

Table 12: Back-of-the-envelope employment effects based on the average reduction in hours estimated via the bunching design and Equation (6), as a function of the demand elasticity for labor (rather than capital) η, and labor supply elasticity α. The bold entry reflects "best-guess" values of η and α.

$\alpha = 0.1$, 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.

D 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 conterfactual 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 $(z_{Bi}, \mathbf{x}_{Bi}) = argmax_{z,\mathbf{x}}\{u_i(z, \mathbf{x}) : z \geq B(\mathbf{x})\}$ where $u_i(z, \mathbf{x})$ is:*

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

and satisfies $u_i(\theta z + (1-\theta)z^*, \theta\mathbf{x} + (1-\theta)\mathbf{x}^*) > \min\{u_i(z,\mathbf{x}), u_i(z^*,\mathbf{x}^*)\}$ *for any* $\theta \in (0,1)$ *and points* $(z,\mathbf{x})$,$(z^*,\mathbf{x}^*)$ *such that* $y_i(\mathbf{x}) \neq k$ *and* $y_i(\mathbf{x}^*) \neq k$.

For generality, I here use weaker notion of convexity of preferences from Assumption CONVEX in Appendix A. It is implied by strict quasi-concavity of $u_i(z,\mathbf{x})$.

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 $B(h)$:

$$h = \underset{h}{\operatorname{argmax}}\, \beta\,(f(h) - z) + (1-\beta)(z - \nu(h)) \quad \textit{with} \quad z = B(h) \tag{7}$$

where $f(h) - z$ is firm profits with concave production f, $z - \nu(h)$ 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 z). Rearranging the maximand of Equation (7) as $(1-2\beta)z + \{\beta f(h) - (1-\beta)\nu(h)\}$, 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 pay z, 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 the more general CONVEX* holds regardless of the distribution of β over worker-firm pairs. If $\beta_{it} < 1/2$, paycheck *it* will look like a worker-chooser, and if $\beta_{it} > 1/2$ paycheck *it* will look 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: intuitively, if a worker is willing to work 40 hours then they will also find it worthwhile to work more, given the sudden increase in their wage. Indeed under regularity conditions all of the data local to 40 are from firm-choosers (or counterfactual bunchers). To make this claim precise, assume that for worker-choosers, hours are the only margin of response (i.e. their utility depends on $\mathbf{x}$ only thought $y(\mathbf{x})$), 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, and let $M_{it} := \sup_y \max\{|IC''_{0it}(y)|, |IC''_{1it}(y)|\}$, where 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*:*

- $P(h_{it} = k$ *and* $K^*_{it} = 0) = P(h_{1it} \leq k \leq h_{0it}$ *and* $K^*_{it} = 0$ *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 Appendix H. □

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 on M_{it}/w_{it}, 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:*

1. $P(h_{0it}=k)=P(h_{1it}=k)=p$
2. $Y_{it}=k$ *iff* ($h_{0it}\in[k,k+\Delta^*_0]$ *and* $W_{it}=0$) *iff* ($h_{1it}\in[k-\Delta^*_1,k]$ *and* $W_{it}=0$), *for some* Δ^*_0,Δ^*_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:*

$$\mathbb{E}[h_{0it}-h_{1it}|h_{it}=k, K^*_{it}=0]\in\left[\Delta^L_k,\Delta^U_k\right]$$

where

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

and

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

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 Appendix H. □

Identification with worker-choosers

Theorem 1* does not immediately yield identification of the buncher-ATE bounds Δ_k^L and Δ_k^U, as we need to estimate each of the arguments to the function g. As shown in the proof of Theorem 1*, 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 $P(W_{it} = 0) \lim_{y\uparrow k} f_{1|W=0}(y)$, and two unidentified probabilities: $P(h_{0it} < k$ and $h_{it} = h_{0it}$ and $W_{it} = 1)$ and $P(Y_{1it} > k$ and $h_{it} = h_{1it}$ and $W_{it} = 1)$.

To illustrate the unidentified quantities, Figure 15 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) dots south of the 45-degree line representing firm-choosers, and the (blue and orange) points above representing 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.

A distribution with worker-choosers

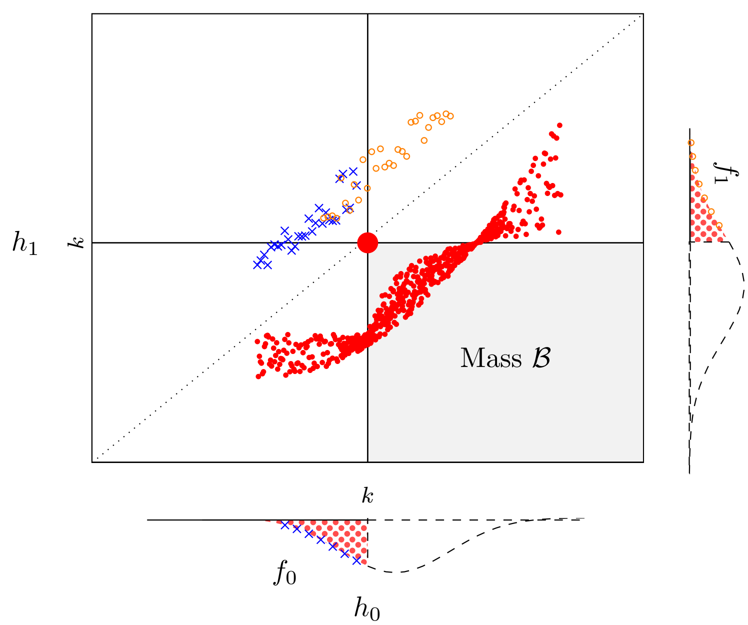


Figure 15: The joint distribution of (h_{0it}, h_{1it}), for a distribution including worker-choosers and satisfying assumption RANK*, cf. Figure 5. See text for description.

Observed to the 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 $P(h_{it} \leq h)$ for $h < 40$ doesn't exactly identify $P(h_{0it} \leq h)$ because some blue 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 thoug they have $h_0 < 40$. Similarly, some orange circles do not appear in the observations 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 15. However, in practice this has little impact on the results, as the bounds Δ_k^L and Δ_k^U are not very sensitive to the values of the CDF inputs $F_{0|W=0,K^*=0}(k)$ and $F_{1|W=0,K^*=0}(k)$. The bounds mostly depend on the density estimates and the size of the bunching mass, given their empirical values. Thus Theorem 1* still partially identifies the buncher LATE among firm-choosers, to a good approximation.

However, a further caveat of Theorem 1* is worth mentioning. 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. However, the survey evidence mentioned in Section 2 suggests that the set of worker-choosers is relatively small, mitigating this concern.

E Interdependencies among hours within the firm

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

The basic issue is as follows: when a single firm chooses hours jointly among mulitple 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. In my potential outcomes donation, this represents a violation of the non-interference condition of the stable unit treatment value assumption (SUTVA), when using the treatment effects Δ_{it} to assess the average impact of the FLSA on hours. If firms maximize profits given a production function that is not linearly separable across workers or across weeks, such violations may occur.

To simplify the notation, suppose that 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 $\mathtt{h}_{0i}(\cdot)$ 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

$\mathtt{h}_{1i}(\cdot)$ analogously with $1.5w_i$. In this notation, the potential outcomes defined in Section 4 are $h_{0i} = \mathtt{h}_{0i}(\mathbf{h_{-i}})$ and $h_{1i} = \mathtt{h}_{1i}(\mathbf{h_{-i}})$. 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 $\mathbb{E}[h_i - h_i^*]$.

Even if there are SUTVA violations, treatment effects $\Delta_i = \mathtt{h}_{0i}(\mathbf{h_{-i}}) - \mathtt{h}_{1i}(\mathbf{h_{-i}})$ remain meaningful as a reduced-form average labor demand elasticity, in which the wage of worker i is changed but with $\mathbf{h_{-i}}$ held fixed. Furthermore, bunching is still informative about identify the buncher ATE: bunching will not occur unless $\Delta_i > 0$ from some units near the kink such that $h_{0i} \in [k, k + \Delta_i]$. The question is whether the treatment effects Δ remain a good guide to the overall effect of the FLSA, given that it may also change $\mathbf{h_{-i}}$ for a given worker i.

For concreteness, let us now 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 $(h_i, \mathbf{h}_{-i}) = \text{argmax}\left\{F(\mathbf{h}) - \sum_j B_{kj}(h_j)\right\}$, where the sum is across workers j and $B_{kj}(h) := w_j h + .5 w_j \mathbb{1}(h > 40)(h - 40)$. Similarly $(h_i^*, \mathbf{h}^*_{-i}) = \text{argmax}\left\{F(\mathbf{h}) - \sum_j w_j h_j\right\}$. Whether $\mathtt{h}_{0i}(\mathbf{h_{-i}})$ 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:

$$F(\mathbf{h}) = \prod_{m=1}^{M} \left((\sum_{i \in \theta_m} a_i \cdot h_i^{\rho_m})^{1/\rho_m} \right)^{\alpha_m} \tag{8}$$

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 ρ_m, which then combine in a Cobb-Douglas form across occupations with exponents α_m. For this production function, 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 θ_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 overall contribution of interdependencies on our parameter of interest θ (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. (8). I therefore turn to an indirect empirical test of whether these effects are likely to play a significant role in θ.

An ideal test of interdependencies between hours within a firm would leverage random individual-level shocks to a worker's hours, and look for a response in the hours of their colleagues. A worker taking sick-pay—thus reducing their hours of work—represents a compelling candidate as its timing may be uncorrelated with that of firm-level shocks (after controlling for seasonality). Figure 16 uses an event study design to show 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). Figure 17 produces similar results when repacing the two-wage-fixed specification of Figure 16 with an "imputation"-based approach similar to Borusyak et al. (2021) and Gardner (2021).

Table 13 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. However, in Table 13 below, the confidence intervals for all three firm size categories overlap, 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	[-1.525, -1.455]	0.027	[-0.231, -0.171]
	[0.189, 0.208]	[-1.676, -1.299]	[0.023, 0.031]	[-0.274, -0.139]
Medium firms	0.103	[-1.123, -0.786]	0.030	[-0.337, -0.224]
	[0.095, 0.110]	[-1.237, -0.710]	[0.025, 0.035]	[-0.407, -0.178]
Large firms	0.050	[-0.768, -0.468]	0.024	[-0.371, -0.224]
	[0.047, 0.054]	[-0.861, -0.414]	[0.021, 0.028]	[-0.444, -0.180]

Table 13: 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.

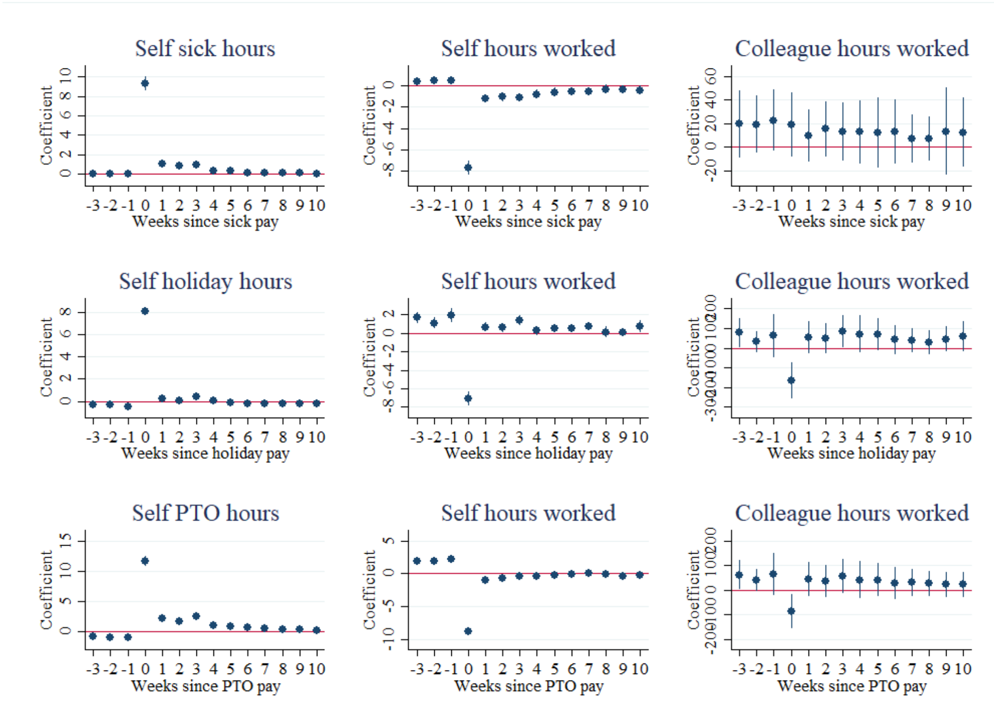


Figure 16: Event study coefficients β_j and 95% confidence intervals across an instance of a worker receiving pay for non-work hours (either sick pay, holiday pay, or paid time off–'PTO'). Confidence intervals are constructed by non-parametric bootstrap clustered by firm. Estimating equation is $y_{it} = \mu_t + \lambda_i + \sum_{j=-3}^{10} \beta_j D_{it,j} + u_{it}$, 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), λ_i are worker fixed effects, and μ_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 both 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). Meanwhile 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.

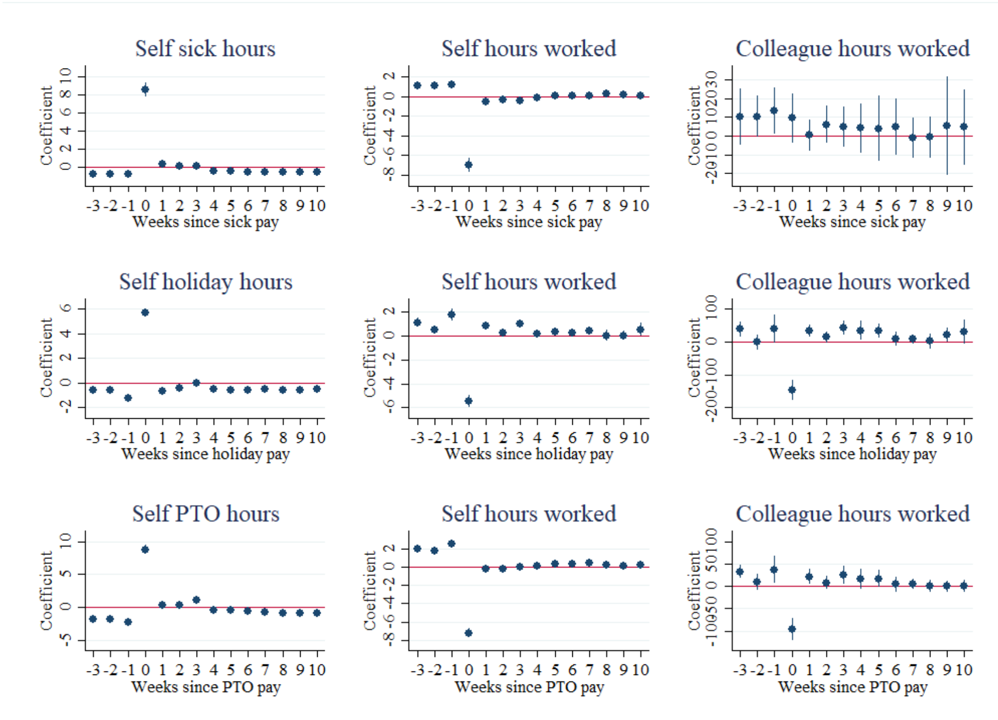


Figure 17: This figure replaces the two-way-fixed-effects estimator used in Figure 16 with an "imputation" approach similar to Borusyak et al. (2021) and Gardner (2021). Results are very similar to those in Figure 16. Specifically, I call all observations that are not between 3 weeks before and 10 weeks after a spell of non-work hours "clean controls", and estimate a first regression $y_{it} = \mu_t + \lambda_i + \epsilon_{it}$ using these observations only. This regression includes all paychecks for workers that never have the corresponding type of non-work hour (sick pay, holiday pay, or PTO), but also a subset of paychecks for nearly all workers who do have a spell of non-work hours at some point (allowing me to estimate their fixed effect λ_i). Given the $\hat{\mu}_t$ and $\hat{\lambda}_i$, I compute $\tilde{y}_{it} = y_{it} - \hat{\mu}_t - \hat{\lambda}_i$ among units that are not clean controls (i.e. those between -3 and 10 weeks after the start of a spell), and estimate a second regression $\tilde{y}_{it} = \sum_{j=-3}^{10} \beta_j D_{it,j} + e_{it}$ on these units only (dropping a small number of workers i for whom there were no clean-control observations). 95% confidence intervals are constructed by non-parametric bootstrap clustered by firm.

F Modeling the determination of wages and "typical" hours

F.1 A simple model with exogenous labor supply

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:

$$\min_{z,h,K,N} N \cdot (z + \psi) + rK \;\text{ s.t. }\; F(Ne(h), K) \geq Q \text{ and } N \leq N(z, h) \tag{9}$$

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 ψ represents non-wage costs per worker. The quantity ψ 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 K. For simplicity, workers within a firm are here identical and all covered by the FLSA.

To understand the properties of the solution to Equation (9), let us examine two illustrative special cases.

Special case 1: an exogenous competitive straight-time wage (the "fixed-wage model")

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 (9) 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, where the function $w_s(\cdot)$ is defined in Equation (1). Then Equation (9) reduces to:

$$\min_{N,h,K} N \cdot (hw + \mathbb{1}(h > 40)(w/2)(h-40) + \psi) + rK \text{ s.t. } F(Ne(h), K) \geq Q \tag{10}$$

By limiting the scope of labor supply effects in the firm's decision, Equation (10) is well-suited to illustrating the competing forces that shape hours choice on the production side: namely the fixed costs ψ on the one hand and the concavity of $e(h)$ on the other. Were ψ equal to zero with $e(h)$ strictly concave globally, a firm solving Equation (10) would always find it cheaper to produce a given level of output with more workers working less hours each. On the other hand, were ψ 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.[15]

Equation (10) 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 Appendix F.2.

Special case 2: iso-elastic functional forms (the "fixed-job model")

By placing some functional form restrictions on Equation (9), we can obtain a closed-form expression for (z^*, h^*). In particular, when both labor supply and $e(h)$ are iso-elastic, production is separable between capital and labor and linear in the latter, and firms set the

[15]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.

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 ϕ; and iii) Q is chosen to maximize profits, the (z^*, h^*) that solve Equation (9) are:*

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

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 β_z.

Proof. See Appendix H. □

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 relative to earnings. The more sensitive labor supply is to a marginal increase in hours as compared with compensation, the higher β will be and lower the optimal number of hours. The proof of Proposition 4 also shows that the general model with $N(z,h)$ differentiable (unlike in Special Case 1) can support an interior solution for hours even without fixed costs $\psi = 0$. Proposition 4 provides an example of the *fixed-job* model: in the absence of perfect competition on the straight-wage, anticipated hours h^*, total pay z^*, and employment $N^* := N_0 \cdot (z^*)^{\beta_z}(h^*)^{\beta_h}$ are unaffected by the FLSA overtime rule, in this simple static model.

F.2 Endogenizing labor supply in an equilibrium search model

The last section treated the labor supply function $N(z,h)$ as exogenous, but in general it might be viewed as an equilibrium object that reflects both worker preferences over income/leisure and the competitive environment for labor. It is conceivable that equilibrium forces would lead to a labor supply function like that of the fixed-wage model, in which the FLSA has an effect on the hours set at hiring.

In this section, I show that the prediction of the fixed-job model that the FLSA has litte to no effect on h^* or z^* is robust to embedding Equation (9) into an extension of the Burdett and Mortensen (1998) model of equilibrium with on-the-job search.[16] 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

[16] 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.

hours or straight-time wages. This motivates the strategy in Section 4.4, in which z^* and h^* are treated as fixed when considering the impact of the FLSA on straight-wages.

F.2.1 The model

I focus on a minimal extension of Burdett and Mortensen (1998) that takes firms to be homogeneous in their technology and workers to be homogeneous in their tastes over the tradeoff between income and working hours.[17] Let there be a large number N_w of workers and large number N_f of firms, and define $m = N_w/N_f$.[18] Formally, we model this as a continuum of workers with mass m, and continuum of firms with unit mass. Firms choose a value of pay z and hours h to apply to all of their workers. Each period, there is an exogenous probability λ that any given worker receives a job offer, drawn uniformly from the set of all firms. Employed workers accept a job offer when they receive an earnings-hours package that they prefer to the one they currently hold, where preferences are captured by a utility function $u(z, h)$ taken to be homogeneous across workers and strictly quasiconcave, where $u_z > 0$ and $u_h < 0$. If a worker is not currently employed, they leave unemployment for a job offer if $u(z, h) \geq u(b, 0)$, where b represents a reservation earnings level required to incent a worker to enter employment. Workers leave the labor market with probability δ each period, and an equal number enters the non-employed labor force.

Before we turn to earnings-hours posting decision of firms, we can already derive several relationships that must hold for the earnings-hours distribution in a steady state equilibrium. First note that the share unemployed v of the workforce must be $v = \frac{\delta}{\delta+\lambda}$, since mass $m(1-v)\delta$ enters unemployment each period, and $m\lambda v$ leaves (taking for granted here that firms only post job offers that are preferred to unemployment, which is indeed a feature of the actual equilibrium). Let's say that job (z, h) is "inferior" to (z', h') when $u(z, h) \leq u(z', h')$. Let P_{ZH} be the firm-level distribution over offers (Z_j, H_j), and define

$$F(z, h) := P_{ZH}(u(Z_j, H_j) \leq u(z, h)) \tag{11}$$

to be the fraction of firms offering inferior job packages to (z, h). The separation rate of workers at a firm choosing (z, h) is thus: $s(z, h) = \delta+\lambda(1-F(z, h))$. To derive the recruitment of new workers to a given firm each period, we define the related quantity $G(z, h)$ – the fraction of employed workers that are at inferior firms to (z, h). In a steady state, note that $G(z, h)$

[17]The model presented here bears similarity to that of Hwang et al. (1998), which also considers search equilibrium with non-wage amenities such as hours. My model generalizes the preferences of workers to be possibly non-quasilinear, which allows my model to support hours dispersion in equilibrium, even with identical firms. In their model, by contrast, firms are allowed to be heterogeneous but all firms with the same production technology would offer the same quantity of hours.

[18]Here we largely follow the notation of the presentation of the Burdett & Mortensen model by Manning (2003).

must satisfy

$$\underbrace{m(1-v)\cdot G(z,h)(\delta+\lambda(1-F(z,h))}_{\text{mass of workers leaving set of inferior firms}} = \underbrace{mv\lambda F(z,h)}_{\text{mass of workers entering set of inferior firms}}$$

since the number of workers at firms inferior to (z,h) is assumed to stay constant. To get the RHS of the above, note that workers only enter the set of firms inferior to (z,h) from unemployment, and not from firms that they prefer. This expression allows us to obtain the recruitment function $R(z,h)$ to a firm offering (z,h). Recruits will come from inferior firms and from unemployment, so that

$$\begin{aligned}
R(z,h) &= \lambda m\left((1-v)G(z,h)+v\right) \\
&= \lambda mv\left(\frac{\lambda F(z,h)}{\delta+\lambda(1-F(z,h))}+1\right) \\
&= m\left(\frac{\delta\lambda}{\delta+\lambda(1-F(z,h))}\right)
\end{aligned}$$

Combining with the separation rate, we obtain the steady-state labor supply function facing each firm:

$$N(z,h) = R(z,h)/s(z,h) = \frac{m\delta\lambda}{\left(\delta+\lambda(1-F(z,h)\right)^2} \tag{12}$$

Eq. (12) is analogous to the baseline Burdett and Mortensen model without hours, with the quantity $F(z,h)$ playing the role of the firm-level CDF of wages from the baseline model.

Now we turn to how the form of $F(z,h)$ in general equilibrium. We take the profits of firms to be

$$\pi(z,h) = N(z,h)(p(h)-z) = m\delta\lambda\cdot\frac{p(h)-z}{\left(\delta+\lambda(1-F(z,h)\right)^2} \tag{13}$$

where the function $p(h)$ corresponds to net revenue per worker $e(h)-\psi$, with $e(h)$ being a weakly concave and increasing "effective labor" function with $e(0)=0$, and z recurring non-wage costs per worker. To simplify some of the exposition, we will emphasize the simplest case of $p(h)=p\cdot h$, such that worker hours are perfectly substitutable across workers.

In equilibrium, the identical firms each playing a best response to $F(z,h)$, and thus all choices of (z,h) in the support of P_{ZH} must yield the same level of profits π^*. This gives an expression for $F(z,h)$ over all (z,h) in the support of P_{ZH}, in terms of π^*:

$$F(z,h) = 1+\frac{\delta}{\lambda}-\sqrt{\frac{m\delta}{\lambda}\cdot\frac{p(h)-z}{\pi^*}} \tag{14}$$

where we subtract the positive square root since the negative square root cannot deliver a real number less than or equal to unity for $F(z,h)$. Note that Eq. (14) only needs to hold at (z,h) that are actually chosen by firms in equilibrium

It follows from Eqs. (14) and (12) that we can rank firms in equilibrium by $F(z,h)$ and therefore by size according to the quantity $z-p(h)$. Note that since Eq. (12) is continuously differentiable in (z,h), we can rule out mass points in P_{ZH} by an argument paralleling that in Burdett and Mortensen (1998). Suppose $P_{ZH}(z,h)=\delta>0$ for some (z,h). Then any firm located at (z,h) and earning positive profits could increase their profits further by offering a sufficiently small increase in compensation (or reduction in hours, or a combination of both). Since $F(z+\delta_z,h)=F(z,h)+\delta$ to first order, there exists a small enough δ_z such that $\pi(z+\delta_t,h)>\pi(z,h)$ by Eq. (13).

To fully characterize the equilibrium P_{ZH}, I first argue that for a strictly quasiconcave utility function u, workers cannot be indifferent between more than two points (z,h) that share a value of $z-p(h)$ (see Figure 18 below). This implies that offers in the support of P_{ZH} lie along a one dimensional path through $\mathbb{R}^2$. Consider for example the case of perfect hours substitutability: $p(h)=ph$, and imagine moving from a given point (z,h) in the support of P_{ZH} continuously along a line that that keeps $z-ph$ and hence $F(z,h)$ constant. Since $F(z,h)$ is constant along this line, we must have that either worker utility is constant or that P_{ZH} has no additional mass along the line. However, we cannot be moving along an indifference curve of $u(z,h)$, as strict convexity of preferences implies that the marginal rate of substitution between compensation and hours can equal p (or more generally $p'(h)$, which is non-increasing) at no more than a single point for a single level of utility. Thus, P_{ZH} puts a positive density on at most one point along each isoquant of $z-p(h)$, and must have positive density on each isoquant within some connected interval. Given this, we can parametrize the points in support of P_{ZH} by a single scalar $t\in[0,1]$, such that $\text{supp}(P_{ZH})=\{(z(t),h(t))\}_{t\in[0,1]}$ and $t=F(z(t),h(t))$.

Now observe that each $(z(t),h(t))$ must pick out the point along its respective isoquant of $z-p(h)$ which delivers the highest utility to workers, i.e.:

$$(z(t),h(t))=\text{argmax}_{z,h}u(z,h) \text{ s.t. } z-p(h)=\eta(t) \tag{15}$$

where $\eta(t)=\frac{\pi^*\lambda}{m\delta}(1-\frac{t}{1+\delta/\lambda})^2$ is the value of $p(h(t))-z(t)$ such that $F(z(t),h(t))=t$ according to Eq.(14), viewed as a function of t. If instead we had $u(z(t),h(t))<\max_{(z,h):z-p(h)=F^{-1}(t)}u(z,h)$, then any firm located at $(z(t),h(t))$ could profitably deviate to the argmax while keeping profits per worker constant but increasing their labor supply by attracting workers from $(z(t),h(t))$. The first order condition for this problem implies that $(z(t),h(t))$ maintains a marginal rate of substitution of $p'(h(t))$ (p in the baseline case) between compensation and hours at all t, while the slope of the path $(z(t),h(t))$ can be derived from the implicit function theorem:

$$\frac{z'(t)}{h'(t)}=-\frac{u_{hh}(z,h)+p''(h)u_z(z,h)+p'(h)u_{zh}(z,h)}{p'(h)u_{zz}(z,h)+u_{zh}(z,h)}\bigg|_{(z,h)=(z(t),h(t))}$$

The curve AB shown in Figure 18 depicts the path $\{(z(t),h(t))\}_{t\in[0,1]}$ for a case in which

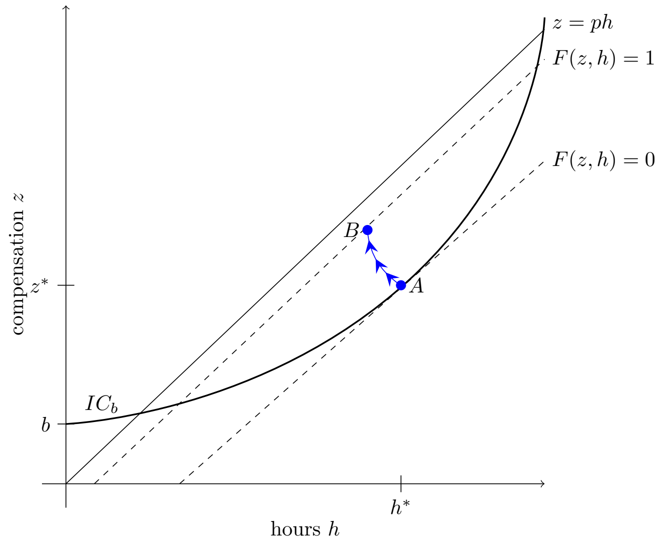


Figure 18: The support of the equilibrium distribution of compensation-hours offers (z, h) lies along the arrowed (blue) curve AB. Figure shows the case of perfect hours substitutability $p(h) = ph$. Plain curve IC_b is the indifference curve passing through the unemployment point $(b, 0)$. The least desirable firm in the economy lies at the pair (z^*, h^*), labeled by A, where IC_b has a slope of p. The other points chosen by firms are found by starting at point A and moving in the direction of higher utility, while maintaining a marginal rate of substitution of p between hours and earnings. This path intersects the line of solutions to $F(z, h) = 1$ given Eq. (14) at point B. Note that this line still lies below the zero profit line $z = ph$, as firms make positive profit. Curve AB is shown for a general non-quasilinear, non-homothetic utility function (see text for details).

preferences are neither homothetic nor quasilinear, for example: $u(z,h) = \frac{z^{1-\gamma}}{1-\gamma} - \beta\frac{h^{1+1/\epsilon}}{1+1/\epsilon}$. If preferences were instead homothetic then AB would be a straight line pointing to the northwest from A. In the numerical calibration, I take preferences to follow the non-quasilinear Stone-Geary functional form.[19] If preferences were quasilinear in income (for example the above with $\gamma = 0$), then AB would be a vertical line rising from point A and there would be no hours dispersion in equilibrium (as in Hwang et al., 1998).

To pin down the initial point A, we note that it must lie on the indifference curve passing through the unemployment point $(b, 0)$, labeled as IC_b in Figure 18. If it were to the northwest of the IC_b curve, a firm located there could increase profits by offering a lower value of $z - p(h)$, since given that $F(z(0), h(0)) = 0$ their steady state labor supply already only recruits from unemployment. However, they cannot offer a pair that lies to the southeast of IC_b, since they could never attract workers from unemployment. I assume that the marginal rate of substitution between compensation and hours is less than $p'(0)$ at $(z, h) = (b, 0)$ (such that there are gains from trade) and increases continuously with h, eventually passing $p'(h)$ at

[19]A CES generalization of Stone-Geary preferences also results in a straight line AB: $u(z,h) = [\theta(z-\gamma_z)^\lambda + (1-\theta)(\gamma_h - h)\lambda]^{1/\lambda}$. It is also possible to obtain a non-linear path AB while maintaining constant elasticity of substitution between earnings and leisure. The work of Sato (1975) on production functions suggests utility functions satisfying $\frac{u_z(z,h)}{u_h(z,h)} = \left(\frac{z-\gamma_z}{h-\gamma_h}\right)^{\frac{1}{1-\lambda}} \phi\left(u(c,h)\right)$ where ϕ is any positive function.

some point h^*. This point is unique given strict quasiconcavity of $u(\cdot)$. Then, let z^* be the earnings value such that workers are indifferent between (z^*, h^*) and unemployment $(b, 0)$, which represents a reservation level of utility required to enter employment.

Finally, we can also express $F(z, h)$ as a function of $(z^*, h^*) = (z(0), h(0))$. Using that $F(z^*, h^*) = 0$ and $\pi^* = \pi(z^*, h^*)$, we can rewrite Equation (14) as:

$$F(z,h) = \left(1 + \frac{\delta}{\lambda}\right)\left[1 - \sqrt{\frac{p(h) - z}{p(h^*) - z^*}}\right] \tag{16}$$

The firms at point B in Figure 18 thus solve $z - p(h) = \left(\frac{\delta}{\delta+\lambda}\right)^2 (z^* - p(h^*))$. Equilibrium profits are $\pi^* = m(p(h^*) - z^*) \cdot \frac{\lambda/\delta}{(1+\lambda/\delta)^2}$. By Eq. (16) we can also work out that $\eta(t) = \left(1 - \frac{t}{1+\delta/\lambda}\right)^2 (ph^* - z^* - \psi)$. Note that in equilibrium, there exists dispersion not only in both earnings and in hours (provided preferences are not quasi-linear), but also in effective hourly wages, as the ratio $z(t)/h(t)$ is also strictly increasing with t. Note that π^* goes to zero in the limit that $\lambda/\delta \to \infty$. In this limit dispersion over hours, earnings, and hourly earnings all disappear as the line AB collapses to a single point on the zero profit line $z = p(h)$.[20]

F.2.2 Effects of FLSA policies

Now consider the introduction of a minimum wage, which introduces a floor on the hourly wage $w := z/h$. I assume that the point (z^*, h^*) does not satisfy the minimum wage, so that the minimum wage binds and rules out part of the unregulated support of P_{ZH}. The left panel of Figure 19 depicts the resulting equilibrium, in which the initial point $(z(0), h(0))$ moves to the point marked A', at which the marginal rate of substitution between compensation and hours is $p'(h)$, but the compensation-hours pair just meets the minimum wage. This compresses the distribution P_{ZH} compared with the unregulated equilibrium from Figure 18, which now follows a subset of the original path AB. In a stochastic dominance sense, all jobs see a reduction in hours and an increase in total compensation (and hence a compounded effect on hourly wages) when a minimum wage is introduced or increased.

The right panel of Figure 19 shows how equilibrium is further affected if in addition to a binding minimum wage, premium pay is required at a higher minimum wage $1.5\underline{\text{w}}$ for hours in excess of 40, provided that the point A' lies at an hours value that is greater than 40. In this case, $(z(0), h(0))$ will lie at point A'', at which the marginal rate of substitution between compensation and hours is equal to h', and compensation is equal to the minimum-compensation function under both the minimum wage and overtime policies: $\underline{\text{w}}(h) := \underline{\text{w}}h + 1(h > 40)(h - 40)\underline{\text{w}}/2$.

[20]Note that there is no contradiction here as the argument that point A must be on IC_b relies on $F(z(0), h(0)) = 0$, which is implied by no mass points in P_{ZH}, in turn implied by firms making positive profit.

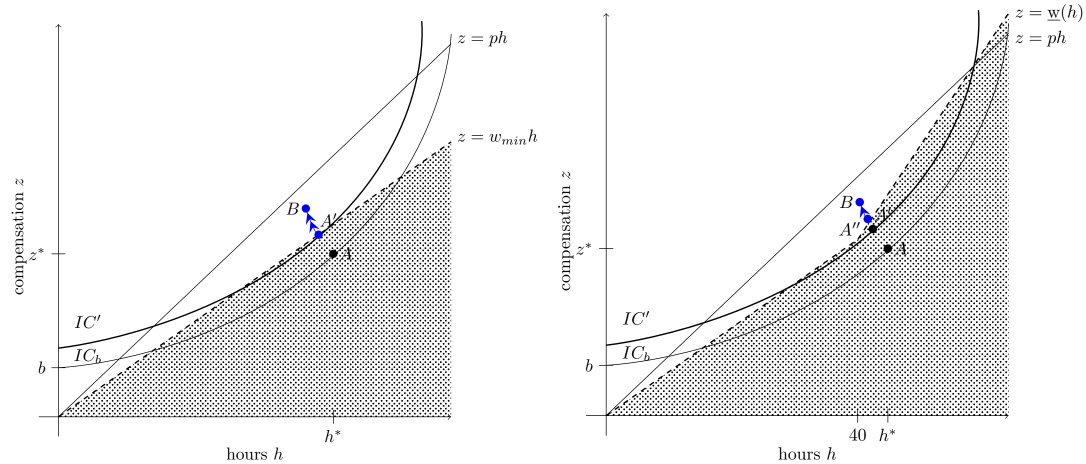


Figure 19: Left panel shows the support of the equilibrium distribution of compensation-hours offers (z, h) under a binding minimum wage. The compensation hours pairs that do not meet $\underline{w}$ are indicated by the shaded region. The lowest-wage offer in the economy moves from point A in the unregulated equilibrium to the point A' on the minimum wage line $z = \underline{w}h$ at which the marginal rate of substitution between compensation and hours equals p. This is equal to the point at which curve AB from Figure 18 crosses the minimum wage line. Curve $A'B$ traces the remainder of curve AB. The compensation-hours offers are thus more compressed and the new distribution of earnings stochastically dominates the distribution from the unregulated equilibrium, while the opposite is true of hours. Right panel shows how this effect is augmented when overtime premium pay for hours in excess of 40 is required, and A' lies at greater than 40 hours. In this case the support of P_{ZH} begins at point A'', which lies on the kinked minimum wage function $\underline{w}(h)$.

F.2.3 Calibration

To allow wealth effects in worker utility while facilitating calibration based on existing empirical studies, we assume worker utility is Stone-Geary:

$$u(z, h) = \beta \log(z - \gamma_z) + (1 - \beta) \log(\gamma_h - h)$$

This simple specification allows a closed form solution to the path $(z(t), h(t))$, given by the following Proposition, which follows directly from the optimization problem (15), while also working out the initial point $(z(0), h(0))$ in each policy regime. Using this, I can then calibrate the model to consider the effects of FLSA policies on earnings and hours.

Proposition. *Under Stone-Geary preferences and linear production $p(h) = ph - \psi$, the equilibrium offer distribution is a uniform distribution over $\{(z(t), h(t))\}_{t\in[0,1]}$, where:*

$$\begin{pmatrix} z(t) \\ h(t) \end{pmatrix} = \begin{pmatrix} p\beta\gamma_h + (1-\beta)\gamma_z - \beta\psi - \beta\eta(t) \\ \beta\gamma_h + \frac{1-\beta}{p}(\gamma_z + \psi) + \frac{(1-\beta)}{p}\eta(t) \end{pmatrix}$$

where $\eta(t) = \left(1 - \frac{t}{1+\delta/\lambda}\right)^2 \cdot (ph(0) - z(0) - \psi)$. The initial point $(z(0), h(0))$ is

1. *$h(0) = \gamma_h - \left(\frac{(b-\gamma_c)(1-\beta)}{p\beta}\right)^\beta \gamma_h^{1-\beta}$ and $z(0) = z^* = \gamma_z + \left(\frac{p\beta\gamma_h}{1-\beta}\right)^{1-\beta} ((b - \gamma_c)(1 - \beta))^\beta$ in the unregulated equilibrium*

2. $h(0) = (\frac{p\beta}{1-\beta}\gamma_h + \gamma_z)(\underline{w} - \frac{p\beta}{1-\beta})^{-1}$ *and* $z(0) = \underline{w}h(0)$ *with a binding minimum wage of* $\underline{w}$ *(binding in the sense that* $z^* < \underline{w}h^*$*)*

3. $h(0) = (\frac{p\beta}{1-\beta}\gamma_h + \gamma_z + 20\underline{w})(1.5\underline{w} - \frac{p\beta}{1-\beta})^{-1}$ *and* $z(0) = 1.5\underline{w}h(0) - 20\underline{w}$ *with a minimum wage of* $\underline{w}$ *and time-and-a-half overtime pay after 40 hours, if the expression for* $h(0)$ *in item 2. is greater than 40*

Moments with respect to the worker distribution can be evaluated for any measurable function $\phi(z, h)$ *as:*

$$E_{workers}[\phi(Z_i, H_i)] = \left(1 + \frac{\lambda}{\delta}\right) \int_0^1 \phi(z(t), h(t)) \cdot \left(1 + \frac{\lambda}{\delta}(1-t)\right)^{-2} dt$$

I calibrate the model focusing on a lower-wage labor market where productivity is a constant $p = \$15$. I allow non-wage costs of $\psi = \$100$ a week, with the value based on estimates of benefit costs in the low-wage labor market.[21] I take $b = \$250$ corresponding to unemployment benefits that can be accrued at zero weekly hours of work.[22]. We calibrate the factor λ/δ using estimates from Manning (2003) using the proportion of recruits from unemployment. Using Manning's estimates from the US in 1990 of about 55% of recruits coming from unemployment, this implies a value of $\lambda/\delta \approx 3$ in the baseline Burdett and Mortensen, 1998 model.

To calibrate the preference parameters, I first pin down β from estimates of the marginal propensity to reduce earnings after random lottery wins (Imbens et al. 2001; Cesarini et al. 2017). Both of these studies report a value in the neighborhood of $\beta = 0.85$. I take a value of $\gamma_z = \$200$ as the level of consumption at which the marginal willingness to work is infinite, and take $\gamma_h = 50$ hours of work per week. I choose this value according to a rule-of-thumb as the average hours among full-time workers in the US (42.5), divided by β.[23]

Given these values, we can compute moments of functions of the joint distribution of compensation and hours using the Proposition and numerical evaluation of the integrals. Table 14 reports worker-level means of hours, weekly compensation, and the hourly wage z/h, as well as employment and profits per worker averaged across the firm distribution. In the unregulated equilibrium, the lowest-compensated workers work about 49 hours a week earning about \$300, while the highest-compensated workers work about 46 hours and earn more than \$550. This equates to a more than doubling of the hourly wage, which is about \$6 for the $t = 0$ workers and over \$12 for the $t = 1$ workers. For each of the first three variables, the mean is much closer to the $t = 1$ value than the $t = 0$ value, which follows

[21] Specifically, I take a benefit cost of \$2.43 per hour worked for the 10th percentile of wages in 2019: BLS ECEC, multiplied by the average weekly hours worked of 42.5 from the 2018 CPS summary, which results in $102.425 \approx 100$.

[22] We use the UI replacement rate for single adults 2 months after unemployment from the OECD. Taking this for individuals at 2/3 of average income (the lowest available in this table), and then use a BLS figure for average income at the 10% percentile of $22,880$, we have $b \approx \$22,880 \cdot 0.6/52.25 = \$263$

[23] Cesarini et al. (2017) point out that when γ_c and no-unearned income, optimal hours choice is $\beta\gamma_h$. By comparison, these authors calibrate γ_h to be about 35 hours in the Swedish labor market.

from the higher-t firms having more employees. The convexity of the labor supply function across values of t is apparent from the firm size row: the largest firm is about 16 times as large as the smallest, while the average firm size is four times larger than the $t = 0$ firms. The final row reports weekly profits per worker: the average worker captures more than half of the employer surplus for the $t = 0$ worker, whose weekly compensation is comparable to the employer's profit for that worker.

	Unregulated equilibrium			$\underline{\text{w}} = 7.25$	$\underline{\text{w}} = 7.25$ & *OT*	$\underline{\text{w}} = 12$ & *OT*
	t=0	t=1	mean	mean	mean	mean
weekly hours	48.85	45.71	46.34	46.18	46.11	45.51
weekly earnings	297.36	564.68	511.22	524.31	530.93	581.78
hourly wage	6.09	12.35	11.06	11.37	11.53	12.78
firm size / smallest	1.00	16.00	4.00	4.00	4.00	4.00
weekly profit per worker	335.46	20.97	146.76	119.81	106.18	1.49

Table 14: Results from the calibration. The parameter $t \in [0, 1]$ indicates firm rank in desirability from the perspective of workers. Means for weekly hours, weekly earnings, and hourly wages are computed with respect to the worker distribution, while firm size and profits per worker is averaged with respect to the firm distribution.

The third column of Table 14 adds a minimum wage set at the current federal rate of \$7.25. This provides a small increase of about 30 cents to the average hourly wage, which now begins at \$7.25 for $t = 0$ rather than \$6.06. Note that the minimum wage provides spillovers by reallocating firm mass up the entire wage ladder, beyond the mechanical effect of increasing the wages of those previously below 7.25. Average hours worked are decreased slightly due to the minimum wage, by about ten minutes per week. As this effect is mediated by a wealth effect in labor supply, we can expect it to be small unless worker preferences deviate significantly from quasi-linearity with respect to income. With $\beta = .85$, this effect is reasonably modest but non-negligible. In the fourth column, we see that the combination of the minimum wage and overtime premium has little effect beyond the direct effect of the minimum wage: hourly earnings increase another 15 cents and hours worked go down by another 0.07. Finally, we see that increasing the minimum wage to \$12 has much larger effects: adding another dollar to average wages and reducing working time by a bit more than half an hour per week. Given the fixed costs assumed in this calibration, a \$12 minimum wage causes employers to run on extremely thin margins for these workers (who have \$15 an hour productivity). However, note that in this model a minimum wage causes neither an increase nor decrease in aggregate non-employment, as this is governed in the steady state only by λ/δ. Thus, the average absolute firm size is unchanged across the policy environments.

G Additional identification results for the bunching design

This section presents several additional sufficient conditions for point or partial identification in the bunching design, beyond Theorem 1 from the main text. In this section, I continue with the notation Y_i rather than h_{it} as in Appendix A. For simplicity, I in this section assume that Y_0 and Y_1 admit a density everywhere so there is no counterfactual bunching at the kink. However, the results here can be applied given a known $p = P(Y_{0i} = Y_{1i} = k)$, as in Section 4.3, by trimming p from the observed bunching and re-normalizing the distribution $F(y)$.

I first consider parametric assumptions when treatment effects are assumed homogeneous, recasting some existing results from the literature into my generalized framework. Then I turn to nonparametric restrictions that also assume homogeneous treatment effects, before stating some results with heterogeneous treatments.

G.1 A generalized notion of homogeneous treatment effects

Recall that in the isoelastic model, treatment effects are homogeneous across units after a log transformation of the choice variable y. In order to formalize and generalize results from the literature that have focused on the isoleastic model, let begin with a generalized notion of homogenous treatment effects. For any strictly increasing and differentiable transformation $G(\cdot)$, let us define for each unit i:

$$\delta_i^G := G(Y_{0i}) - G(Y_{1i})$$

The iso-elastic model common in the bunching-design literature predicts that while Δ_i is heterogeneous across i, δ_i^G is homogeneous when G is taken to be the natural logarithm function. In this case Δ_i^G is proportional to a reduced form elasticity measuring the percentage change in $y_i(\mathbf{x})$ when moving from constraint B_{1i} to B_{0i}. In particular, in the simplest case of a bunching design in which B_0 and B_1 are linear functions of y with slopes ρ_0 and ρ_1 respectively, and utility follows the iso-elastic quasi-linear form of Equation (4), we have:

$$\delta_i^G = \delta := |\epsilon| \cdot \ln(\rho_1/\rho_0)$$

for all units i, when G is taken to be the natural logarithm.

Note that under CHOICE and CONVEX the result of Lemma 1 holds with $G(\cdot)$ applied to each of Y_i, Y_{0i}, and Y_{1i}, since G is strictly increasing. When δ_i^G is homogeneous for some G with common value δ, we thus have that $\mathcal{B} = P\left(G(Y_{0i}) \in [G(k), G(k) + \delta]\right)$ by Proposition 1. Since $G(\cdot)$ is strictly increasing, we can still write the bunching condition in terms of counterfactual "levels" Y_{0i} as

$$\mathcal{B} = P(Y_{0i} \in [k, k + \Delta]) \text{ where } \Delta = G^{-1}\left(G(k) + \delta\right) - k \tag{17}$$

For example, $\Delta = k(e^\delta - 1)$ in the iso-elastic model. The parameter Δ is equal to the parameter Δ_0^* introduced in Section 4.3, since $\delta_i^G = \delta$ implies rank invariance between Y_{0i} and Y_{1i}. Δ can be seen as a pseudo-parameter plays the same role as Δ would in a setup in which we assumed a constant treatment effects in levels $\Delta_i = \Delta$. If it can be pinned down, it will also be possible to identify δ. Nevertheless, it will be important to keep track of the function G when δ_i^G is assumed homogeneous. For instance, homogeneous $\delta_i^G = \delta$ implies that $f_0^G(G(k)+\delta) = f_1^G(G(k))$ but not that $f_0(k+\Delta) = f_1(k)$, where f_d^G is the density of $G(_{di})$ for each $d \in \{0,1\}$.

G.2 Parametric approaches with homogeneous treatment effects

The approach introduced by Saez 2010 assumes that the density $f_0(y)$ is linear on the bunching interval $[k, k+\Delta]$. This affords point-identification of ϵ in an iso-elastic utility model. We can use the notation above to provide the following generalization of this result:

Proposition 5 (identification by linear interpolation, à la Saez 2010). *If $\delta_i^G = \delta$ for some G, $F_1(y)$ and $F_0(y)$ are continuously differentiable, and $f_0(y)$ is linear on the interval $[k, k+\Delta]$, then with CONVEX, CHOICE:*

$$\mathcal{B} \;=\; \frac{1}{2}\left(G^{-1}\left(G(k)+\delta\right) - k\right)\left\{\lim_{y\uparrow k} f(y) + \frac{G'(G^{-1}\left(G(k)+\delta\right))}{G'(k)} \lim_{y\downarrow k} f(y)\right\}$$

Proof. See Section H. □

In particular, given the iso-elastic model with budget slopes ρ_0 and ρ_1:

$$\mathcal{B} = \frac{\Delta}{2}\left\{\lim_{y\uparrow k} f(y) + \frac{k}{k+\Delta}\lim_{y\downarrow k} f(y)\right\} = \frac{k}{2}\left(\left(\frac{\rho_0}{\rho_1}\right)^{\epsilon} - 1\right)\left(\lim_{y\uparrow k} f(y) + \left(\frac{\rho_0}{\rho_1}\right)^{-\epsilon}\lim_{y\downarrow k} f(y)\right) \tag{18}$$

which serves as the main estimating equation from Saez (2010) (and can be solved for ϵ by the quadratic formula). The empirical approach of Saez (2010) can thus be seen as applying a result justified in a much more general model than the iso-elastic utility function assumed therein, provided that the researcher is willing to assume homogeneous treatment effects (possibly after some known transformation G, and/or conditional on observables).[24] Note that the linearity assumption of Proposition 5 could be falsified by visual inspection: it implies that right and left limits of the derivative of the density of Y_i at the kink are equal.

A more popular approach, following Chetty et al. (2011), is to use a global polynomial approximation to $f_0(y)$, which interpolates $f_0(y)$ inwards from both directions across the

[24]Note that if we had instead assumed that $f_0^G(y)$ is linear (on the interval $[G(k), G(k)+\delta^G]$), then we simply replace $f(y)$ by $f^G(y)$ in the above and let G be the identity function, which can be readily solved for δ^G with the simpler expression $\delta^G = \mathcal{B}/\frac{1}{2}\left\{\lim_{y\uparrow k} f^G(y) + \lim_{y\downarrow k} f^G(y)\right\}$.

missing region of unknown width Δ. This technique has the added advantage of accommodating diffuse bunching, for which the relevant $\mathcal{B}$ is the total "excess-mass" around k rather than a perfect point mass at k. I focus here on the simplest case in which bunching is exact, as in the overtime setting. The polynomial approach can be seen as a special case of the following result:

Proposition 6 (identification from global parametric fit, à la Chetty et al. 2011). *Suppose $f_0(y)$ exists and belongs to a parametric family $g(y;\theta)$, where $f_0(y) = g(y;\theta_0)$ for some $\theta_0 \in \Theta$, and that $\delta_i^G = \delta$ for some G and CONVEX and CHOICE hold. Then, if:*

1. *$g(y;\theta)$ is an analytic function of y on the interval $[k, k+\Delta]$ for all $\theta \in \Theta$, and*
2. *$g(y;\theta_0) > 0$ for all $y \in [k, k+\Delta]$,*

it follows that Δ (and hence δ) is identified as $\Delta(\theta_0)$, where for any θ, $\Delta(\theta)$ is the unique Δ such that $\mathcal{B} = \int_k^{k+\Delta} g(y;\theta)dy$, and θ_0 satisfies

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta(\theta_0);\theta_0) & y > k \end{cases} \tag{19}$$

Proof. See Section H. □

The standard approach of fitting a high-order polynomial to $f_0(y)$ can satisfy the assumptions of Proposition 6, since polynomial functions are analytic everywhere. Proposition 6 yields an identification result that can justify an estimation approach similar to one often made in the literature, based on Chetty et al. (2011).[25] However, it requires taking seriously the idea that $f_0(y) = g(y;\theta_0)$, treating the approach as parametric rather than as a series approximation to a nonparametric density $f_0(y)$. This assumption is very strong. Indeed, assuming that $g(y;\theta_0)$ follows a polynomial exactly has even more identifying power than is exploited by Proposition 6. In particular, if we also have that $f_1(y) = g(y;\theta_1)$ then we could use data on either side of the kink to identify by θ_0 and θ_1, which would allow identification of the average treatment effect with complete treatment effect heterogeneity.

G.3 Nonparametric approaches with homogeneous treatment effects

The additional assumptions from the preceding section have allowed for point-identification of causal effects under an assumption of homogenous treatment effects. These assumptions have taken the form of parametric restrictions on the density of counterfactual choices Y_{0i} in the missing region $[k, k+\Delta]$: that this density is linear or more generally fits a parametric family

[25] The estimation technique proposed by Chetty et al. (2011) ignores the shift term $\Delta(\theta)$ in Equation (19), a limitation discussed by Kleven (2016). This is perhaps less problematic in typical settings where the bunching is somewhat diffuse around the kink, in contrast to the overtime setting in which bunching is exact, and the slope of the density is far from zero near 40. A more robust estimation procedure for parametric bunching designs could be based on iterating on Equation (19) after updating $\Delta(\theta)$, until convergence. This presents an interesting topic for future research.

of analytic functions. As has been argued in Blomquist and Newey (2017), these parametric assumptions drive all of the identification, an undesirable feature from the standpoint of robustness to departures from them. I now explore some non-parametric assumptions about $f_0(y)$ that yield bounds on Δ in a model with homogeneous treatment effects.

For example, monotonicity of $f_0(y)$ has been suggested by Blomquist and Newey (2017) as an alternative assumption in the context of the iso-elastic model. A result based on monotonicity that allows heterogeneous treatment effects is presented in Section G.4. However, monotonicity may be restrictive if the density of Y_0 has a mode near the kink point. In this case, local log-concavity of $f_0(y)$ may be a more attractive assumption (concavity or convexity would be another). [26] Note that log-concavity is a stronger version of the bi-log-concavity assumption used in the main text, but still nests many common parametric distributions such as the uniform, normal, exponential extreme value and logistic. For simplicity, this result assumes homogeneous treatment effects in levels (rather than after applying a function G).

Proposition 7 (bounds from log-concavity). *Suppose that $\Delta_i = \Delta$ and that $f_0(y)$ is log-concave in the interval $y \in [k, k+\Delta]$ and continuously differentiable at k and $k+\Delta$. Then, under CONVEX and CHOICE:*

$$\Delta \in [\Delta^L, \Delta^U]$$

where

$$\Delta^U = \mathcal{B} \cdot \frac{\ln(f_+) - \ln(f_-)}{f_+ - f_-} \quad \textit{and} \quad \Delta^L = \left(\frac{f_-}{f'_-} - \frac{f_+}{f'_+}\right) \ln\left(\frac{\mathcal{B} + \frac{f_-^2}{f'_-} - \frac{f_+^2}{f'_+}}{\frac{f_-}{f'_-} - \frac{f_+}{f'_+}}\right) + \frac{f_+}{f'_+} \ln f_+ - \frac{f_-}{f'_-} \ln f_-$$

where $f'_- := \lim_{y\uparrow k} f'(y)$ and $f'_+ := \lim_{y \downarrow k} f'(y)$

Proof. See Figure 20. Derivation of expressions available by request. □

Intuition for Proposition 7 is provided in Figure 20. If $f_0(y)$ is log convex rather than log-concave in the missing region, then the bounds Δ^L and Δ^U can simply be swapped. Or, if we suppose that f_0 is *either* log-concave or log-convex locally: $\Delta \in [\min\{\Delta^U, \Delta^L\}, \max\{\Delta^U, \Delta^L\}]$.

G.4 Alternative identification strategies with heterogeneous treatment effects

An argument made in Saez 2010 and Kleven and Waseem (2013) uses a uniform density assumption to allow heterogeneous treatment in the bunching-design. If a kink is very small, then this might be justified as an approximation given smoothness of $f(\Delta, y)$, since Δ_i will be "small" for all i. Below I state an analog of this result in the generalized bunching design

[26] Log concavity has previously been assumed as a shape restriction in the context of bunching by Diamond and Persson (2016), though to study the effects of manipulation on other variables, rather than for the effect of incentives on the variable being manipulated.

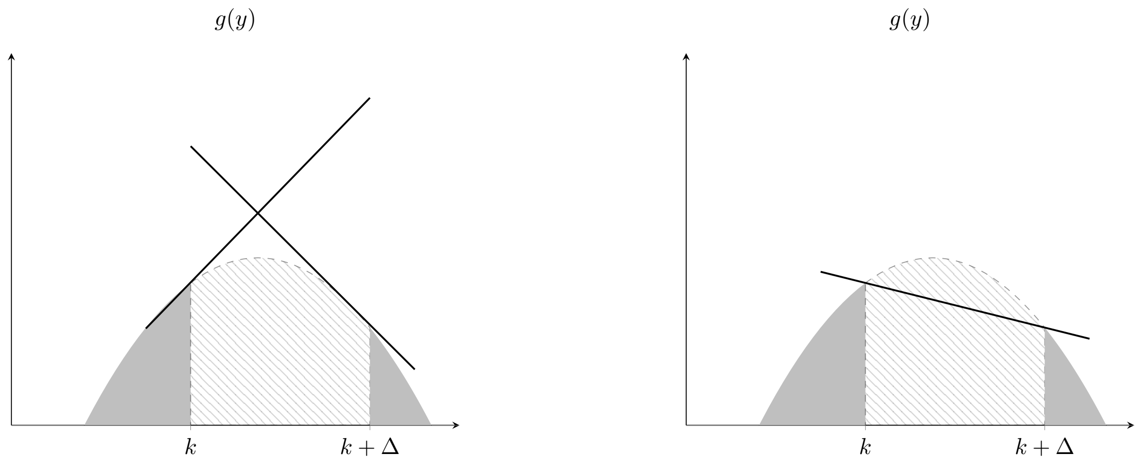


Figure 20: The left and right panels of this figure depict intuition for the lower and upper bounds on Δ in Proposition 7. In both panels, the hatched region is the missing region $[k, k+\Delta]$ which contains known mass $\mathcal{B}$. The function plotted is $g(y)$, the logarithm of $f_0(y)$. Outside of the missing region, this function is known. Concavity of $g(y)$ provides both upper and lower bounds for the values of $g(y)$ inside the missing region, yielding the analytic bounds in Proposition 7.

framework of this paper. The result will make use of the following Lemma, which states that treatment effects must be positive at the kink:

Lemma POS (positive treatment effect at the kink). *Under WARP and CHOICE,* $P(\Delta_i \geq 0 | Y_{0i} = k) = P(\Delta_i \geq 0 | Y_{1i} = k) = 1$.

Proof. See proof of Lemma 1, which rules out the events $Y_{0i} \leq k < Y_{1i}$ and $Y_{0i} < k \leq Y_{1i}$. $\square$

Proposition 8 (identification of an ATE under uniform density approximation). *Let Δ_i and Y_{0i} admit a joint density $f(\Delta, y)$ that is continuous in y at $y = k$. For each Δ assume that $f(\Delta, Y_0) = f(\Delta, k)$ for all Y_0 in the region $[k, k+\Delta]$. Under Assumptions WARP and CHOICE*

$$\mathbb{E}\left[\Delta_i | Y_{0i} = k\right] \geq \frac{\mathcal{B}}{\lim_{y \uparrow k} f(y)},$$

with equality under CONVEX.

Proof. Note that

$$\begin{aligned}
\mathcal{B} \leq P(Y_{0i} \in [k, k+\Delta_i]) &= \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot f(\Delta, y) = \int_0^\infty f(\Delta, k) \Delta d\Delta \\
&= f_0(k) P(\Delta_i \geq 0 | Y_{0i} = k) \mathbb{E}\left[\Delta_i | Y_{0i} = k, \Delta \geq 0\right] \\
&\leq \lim_{y \uparrow k} f(y) \cdot \mathbb{E}\left[\Delta_i | Y_{0i} = k\right]
\end{aligned}$$

using Lemma POS in the last step. The inequalities are equalities under CONVEX. $\square$

Lemma SMALL in Appendix B formalizes the idea that the uniform density approximation from Proposition 8 becomes exact in the limit of a "small" kink.

We can also produce a result based on monotonicity, allowing heterogeneous treatment effects. Let $\tau_0 := E[\Delta_i|Y_{0i} = k]$ and $\tau_1 := E[\Delta_i|Y_{1i} = k]$.

Proposition 9 (monotonicity with heterogeneous treatment effects). *Assume CONVEX and CHOICE, and suppose the joint density $f_0(\Delta, y)$ of Δ_i and Y_{0i} and the joint density $f_1(\Delta, y)$ of Δ_i both exist. Suppose first that $f_0(\Delta, y)$ is weakly increasing on the interval $y \in [k, k + \Delta]$ for all Δ in the support of Δ_i. Then*

$$\tau_1 \geq \frac{\mathcal{B}}{\lim_{y\downarrow k} f(y)} \quad \textit{and} \quad \tau_0 \leq \frac{\mathcal{B}}{\lim_{y\uparrow k} f(y)}$$

Alternatively, if $f_1(\Delta, y)$ is weakly decreasing on the interval $y \in [k - \Delta, k]$ for each Δ, then

$$\tau_0 \geq \frac{\mathcal{B}}{\lim_{y\uparrow k} f(y)} \quad \textit{and} \quad \tau_1 \leq \frac{\mathcal{B}}{\lim_{y\downarrow k} f(y)}$$

Proof. Note that $f_1(\Delta, y) = f_0(\Delta, y + \Delta)$ for any y, Δ, and hence $f_0(y, \Delta)$ is increasing (decreasing) on $[k, k + \Delta]$ whenever $f_1(y, \Delta)$ is increasing (decreasing) on $[k - \Delta, k]$. Then:

$$\begin{aligned}
\mathcal{B} &= \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot f_0(\Delta, y) \leq \int_0^\infty \Delta f_0(\Delta, k) d\Delta = f_0(k)\tau_0 \\
\mathcal{B} &= \int_0^\infty d\Delta \int_{k-\Delta}^{k} dy \cdot f_1(\Delta, y) \geq \int_0^\infty \Delta f_1(\Delta, k) d\Delta = f_1(k)\tau_0
\end{aligned}$$

for example in the first case, where we have used Lemma POS. The reverse case is analogous

□

This result implies that when treatment effects are statistically independent of Y_0 (for example when they are homogenous): $\Delta_i \perp Y_{0i}$, we have that $\mathbb{E}[\Delta_i] = \tau_0 = \tau_1 \in \left[\frac{\mathcal{B}}{\max\{f_-, f_+\}}, \frac{\mathcal{B}}{\min\{f_-, f_+\}}\right]$.

Other approaches to identification with heterogeneous treatment effects are possible when the researcher observes covariates X_i that are unaffected by a counterfactual shift between B_1 and B_0. For example, assuming that $E[X_i|Y_{0i} = y]$ or $E[X_i|Y_{1i} = y]$ are Lipschitz continuous with a known constant leads to a lower bound on maximum of τ_0 and τ_1 from an observed discontinuity of $E[X_i|Y_i = y]$ at $y = k$. Another strategy for using covariates would be to model the potential outcomes Y_{0i} and Y_{1i} as functions of them. If we are willing to suppose that

$$Y_{0i} = g_0(X_i) + U_{0i} \quad \text{and} \quad Y_{1i} = g_1(X_i) + U_{1i}$$

with U_{1i} and U_{0i} each statistically independent of X_i, then the censoring of the distributions of Y_{0i} and Y_{1i} in Lemma 1 can be "undone", following the results of Lewbel and Linton (2002).[27]. This would allow estimation of the unconditional average treatment effect as well as quantile treatment effects at all levels. However, the assumption that U_0 and U_1 are independent of X is quite strong.

[27] Lewbel and Linton (2002) establish identification of $g(x)$ and $F_U(u)$ in a model where the econometrician observes censored observations of $Y = g(X) + U$. Given knowledge of the distribution of X, the estimated marginal distributions of U_1 and U_2, and the estimated function $g(x)$ the researcher could estimate the distributions $F_1(y) = P(Y_{1i} \leq y)$ and $F_0(y) = P(Y_{0i} \leq y)$ by deconvolution, and then estimate causal effects.

G.5 Two bunching design settings from the literature

Below I discuss two examples from the literature that illustrate the general kink bunching design framework described in Section 4. The first is the classic labor supply example, where convexity of preferences arises from increasing opportunity costs of time allocated to labor. In the second example, firms are again the decision makers but now the "running variable" y is a function of two underlying choice variables $\mathbf{x}$.

Example 1: Labor supply with taxation

Individuals have preferences $\tilde{u}_i(c, y)$ over consumption c, and labor earnings y, where ϵ_i is a vector of parameters capturing heterogeneity over the disutility of labor, labor productivity, etc. The agent's budget constraint is $c \leq y - B(y)$ where $B(y)$ is income tax as a function of pre-tax earnings y. $\tilde{u}_i(c, y)$ is taken to be strictly quasi-concave in (c, y) for each i as the opportunity cost of leisure rises with greater earnings, and monotonically increasing in consumption. Define $z = y - c$ to be tax liability, and let $u_i(z, y) = \tilde{u}_i(y - z, y)$ which is monotonically decreasing in tax. Individuals now choose a value of y (e.g. by adjusting hours of work, number of jobs, or misreporting) given a progressive tax schedule $B_k(y) = \tau_0 y + 1(y \geq k)(\tau_1 - \tau_0)(y - k)$, with the kink arising from an increase in marginal tax rates from τ_0 to $\tau_1 > \tau_0$ at $y = k$. The budget functions are $B_0(y) = \tau_0 y$, $B_1(y) = \tau_1 y - (\tau_1 - \tau_0)k$, and the kinked budget constraint can bewritten $z \geq B_k(y) = \max\{B_0(y), B_1(y)\}$.

Example 2: Minimum tax schemes

Best et al. (2015) study a feature of corporate taxation in Pakistan in which firms pay the maximum of a tax on output and a tax on reported profits:

$$B(r, \hat{w}) = \max\{\tau_\pi(r - \hat{w}), \tau_r r\}$$

where r is firm revenue, $\hat{w}$ is reported costs, and $\tau_r < \tau_\pi$. Under the profit tax, firms have incentive to reduce their tax liability by inflating the value $\hat{w}$ above their true costs of production $w_i(r)$. One can write tax liability as a piecewise function in which the tax regime depends on reported profits as a fraction of output: $y = \frac{r - \hat{w}}{r} = 1 - \frac{\hat{w}}{r}$:

$$B(r, \hat{w}) = \begin{cases} \tau_r r & \text{if } y \leq \tau_r/\tau_\pi \\ \tau_\pi(r - \hat{w}) & \text{if } y > \tau_r/\tau_\pi \end{cases}$$

This function has a "kink" in both r and $\hat{w}$ when $y(r, \hat{w}) = k = \tau_r/\tau_\pi$. In this setting, $B_0(r, \hat{w}) = \tau_r r$, corresponding to a tax on output while $B_1(r, \hat{w}) = \tau_\pi(r - \hat{w})$ describes a tax on (reported) profits. Both functions are linear, and hence weakly convex, in the vector $(r, \hat{w})$. The functions B_{0i}, B_{1i} and y_i are all common across firms.

Assume that firm i chooses the pair $\mathbf{x} = (r, \hat{w})$ according to preferences $u_i(z, \mathbf{x})$, which are strictly decreasing in tax liability z and strictly quasiconcave in $(z, r, \hat{w})$. In Best et al. (2015), preferences are for example taken to be in a baseline model:

$$u_i(z, r, \hat{w}) = r - w_i(r) - g_i(\hat{w} - w_i(r)) - z \tag{20}$$

where $g_i(\cdot)$ represents costs of tax evasion by misreporting costs. This specification of $u_i(z, r, \hat{w})$ is strictly quasi-concave provided that the production and evasion cost functions $w_i(\cdot)$ and $g_i(\cdot)$ are strictly convex.

With such preferences, the presence of the minimum tax kink can be expected to lead to a firm response among both margins of $\mathbf{x}$: r and $\hat{w}$. In particular, consider a linear approximation to $\Delta_i = Y_i(0) - Y_i(1)$ for a buncher with $Y_{0i} \approx k$, keeping the i implicit:

$$\begin{aligned}
\Delta &\approx \left.\frac{dy(r,\hat{w})}{\hat{w}}\right|_{(r_0,\hat{w}_0)} \Delta_{\hat{w}} + \left.\frac{dy(r,\hat{w})}{r}\right|_{(r_0,\hat{w}_0)} \Delta_r \\
&= \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(\Delta_{w(r)} + \Delta_{(\hat{w}-w(r))}\right) \\
&\approx \frac{\hat{w}_0}{r_0^2}\Delta_r - \frac{1}{r_0}\left(w'(r_0)\Delta_{ri} + \Delta_{(\hat{w}-w(r))}\right) \\
&= \frac{1}{r_0}\left\{(1 - Y_0 - w'(r_0))\Delta_r \Delta_{\hat{w}}\right\} \approx \frac{1}{r}\left\{-k\Delta_r - \Delta_{(\hat{w}-w)}\right\} \\
&\approx \frac{1}{r_0}\left\{-\frac{\tau_r}{\tau_\pi} \cdot r\epsilon^r \frac{d(1-\tau_E)}{\tau_E} - \Delta_{\hat{w}i}\right\} = \frac{\tau_r^2}{\tau_\pi}\epsilon^r - \frac{\Delta_{(\hat{w}-w)}}{r_0}
\end{aligned} \tag{21}$$

where ϵ^r is the elasticity of firm revenue with respect to the net of effective tax rate $1-\tau_E$. In this case, when crossing from the output to reported profits regime $\frac{d(1-\tau_E)}{\tau_E} = -\tau_r$, implying the final expression (see Best et al. 2015 for definition of τ_E). We have also used the optimality condition that $w'(r_0) = 1$. Expression (21) shows that the response to the minimum tax kink is almost entirely driven by a response on the difference between reported and actual costs: $\hat{w}_i - w_i(r)$. This is because τ_r is less than 1%, so the first term ends up not contributing meaningfully in practice (it scales as the square of τ_r). In this empirical setting, it is thus possible to interpret the bunching response as a response to one of the components of $\mathbf{x}$, despite $\mathbf{x}$ being a vector.

We can also situate the setting of Best et al. (2015) in terms of a continuum of cost functions, as in Section A.6. In particular, let $\rho \in [0, 1]$ and define

$$B(r, \hat{w}; \rho, k) = \frac{\tau_r}{1 - \rho(1-k)}(y - \rho c)$$

Then $B_0(r, \hat{w}) = B(r, \hat{w}; 0)$ and $B_1(r, \hat{w}; \tau_r/\tau_\pi) = B(r, \hat{w}; 1, \tau_r/\tau_\pi)$. It can be verified that for any $\rho' > \rho$ and k, $B(r, \hat{w}; \rho', k) > B(r, \hat{w}; \rho, k)$ iff $y_i(r, \hat{w}) > k$, with equality when $y_i(r, \hat{w}) = k$. The path from $\rho_0 = 0$ to $\rho_1 = 1$ passes through a continuum of tax policies in which the tax base gradually incorporates reported costs, while the tax rate on that tax base also increases continuously with ρ.

H Additional proofs

H.1 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 ($Y_i \geq k$ 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 $\Delta_i = Y_{0i} - Y_{1i}$ 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 $Y_{1i} \in [k - \Delta_i, k]$). 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_1, where we consider the event $Y_{1i} \in [y, y + \delta]$ any $y > k$. Under CONVEX instead of WARP, the inequalities are all equalities, by Lemma 1.

H.2 Proof of Lemma 2

Let $\Delta_i^k(\rho, \rho') := Y_i(\rho, k) - Y_i(\rho', k)$ for any $\rho, \rho' \in [\rho_0, \rho_1]$ and value of k.

Assumption SMOOTH (regularity conditions). *The following hold:*

1. $P(\Delta_i^k(\rho, \rho') \leq \Delta, Y_i(\rho, k) \leq y)$ *is twice continuously differentiable at all* $(\Delta, y) \neq (0, k^*)$, *for any* $\rho, \rho' \in [\rho_0, \rho_1]$ *and* k.
2. $Y_i(\rho, k) = Y(\rho, k, \epsilon_i)$, *where* ϵ_i *has compact support* $E \subset \mathbb{R}^m$ *for some* m. $Y(\cdot, k, \cdot)$ *is continuously differentiable on all of* $[\rho_0, \rho_1] \times E$, *for every* k.
3. *there possibly exists a set* $\mathcal{K}^* \subset E$ *such that* $Y(\rho, k, \epsilon) = k^*$ *for all* $\rho \in [\rho_0, \rho_1]$ *and* $\epsilon \in \mathcal{K}^*$. *The quantity* $\mathbb{E}\left[\frac{\partial Y_i(\rho,k)}{\partial \rho}\middle| Y_i(\rho, k) = y, \epsilon_i \notin \mathcal{K}^*\right]$ *is continuously differentiable in* y *for all* y *including* k^*.

In the remainder of this proof I keep k be implicit in the functions $Y_i(\rho, k)$ and $\Delta_i^k(\rho, \rho')$, as it will remained fixed. Item 1 of SMOOTH excludes the point $(0, k^*)$ on the basis that we may expect point masses at $Y_i(\rho) = k^*$, as in the overtime setting. Following Section 4, item 3 imposes that all such "counterfactual bunchers" have zero treatment effects, while also introducing a further condition that will be used later in Lemma 3. Let K_i^* be an indicator for $\epsilon_i \in \mathcal{K}^*$ and denote $p = P(K_i^* = 1)$. Item 1 implies that the density $f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, y)$ is continuous in y whenever $y \neq k^*$ or $\Delta \neq 0$, so I define $f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k^*) = \lim_{y \to k^*} f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, y)$

for any ρ, ρ' and Δ. Similarly, we can define the marginal density $f_\rho(y)$ of $Y_i(\rho)$ at k^* to be $\lim_{y \to k^*} f_\rho(y)$ for any ρ.

By item 1 of Assumption SMOOTH, the marginal $F_\rho(y) := P(Y_i(\rho) \le y)$ is differentiable away from $y = k$ with derivative $f_\rho(y)$. From the proof of Theorem 1 it follows that $\mathcal{B} \le F_{\rho_1}(k) - F_{\rho_0}(k) + p(k)$ with equality under CONVEX, and thus:

$$\begin{aligned}
\mathcal{B} - p(k) &\le F_{\rho_1}(k) - F_{\rho_0}(k) \\
&= \int_{\rho_0}^{\rho_1} \frac{d}{d\rho} F_\rho(k) d\rho \\
&= \int_{\rho_0}^{\rho_1} \lim_{\delta \downarrow 0} \frac{F_{\rho+\delta}(k) - F_\rho(k)}{\delta} d\rho \\
&= \int_{\rho_1}^{\rho_0} \lim_{\delta \downarrow 0} \frac{P(Y_i(\rho+\delta) \le k \le Y_i(\rho)) - p(k)}{\delta} d\rho \\
&= \int_{\rho_1}^{\rho_0} f_\rho(k) \mathbb{E}\left[-\frac{Y_i(\rho)}{d\rho} \middle| Y_i(\rho) = k \right] d\rho
\end{aligned}$$

where the third equality applies the identity $1 = P(Y_{0i} \le k) + P(Y_i(\rho) \le k \le Y_i(\rho+\delta)) + P(Y_{1i} > k)$ under CHOICE and WARP (this follows from item i) of the proof of Lemma 1) to the pair of choice constraints $B(\rho)$ and $B(\rho+\delta)$, noting that $P(Y_i(\rho) < k) = F_\rho(k) - p(k)$. The final equality uses Lemma SMALL.

H.3 Proof of Lemma SMALL

Throughout this proof we let f_W denote the density of a generic random variable or random vector W_i, if it exists. Write $\Delta_i(\rho, \rho') = \Delta_i(\rho, \rho', \epsilon_i)$ where $\Delta_i(\rho, \rho', \epsilon) := Y(\rho, \epsilon) - Y(\rho', \epsilon)$.

$$\begin{aligned}
\lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \le k \le Y_i(\rho')) - p(k)}{\rho' - \rho} &= \lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \in [k, k + \Delta(\rho, \rho')_i]) - p(k)}{\rho' - \rho} \\
&= \lim_{\rho' \downarrow \rho} \frac{P(Y_i(\rho) \in (k, k + \Delta(\rho, \rho')_i])}{\rho' - \rho} \\
&= \lim_{\rho' \downarrow \rho} \frac{1}{\rho' - \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, y) \\
&= \lim_{\rho' \downarrow \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot \frac{f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k) + (y-k) r_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k, y)}{\rho' - \rho}
\end{aligned} \tag{22}$$

where we have used that by item 1 the joint density of $\Delta_i(\rho, \rho')$ and $Y_i(\rho)$ exists for any ρ, ρ' and is differentiable and $r_{\Delta(\rho,\rho'),Y(\rho)}$ is a first-order Taylor remainder term satisfying

$$\lim_{y \downarrow k} |r_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, y)| = |r_{\Delta(\rho,\rho'),Y(\rho)}(\Delta, k)| = 0$$

for any Δ.

I now show that the whole term corresponding to this remainder is zero. First, note that:

$$\begin{aligned}
\left| \lim_{\rho' \downarrow \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot \frac{(y-k) r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta, y)}{\rho' - \rho} \right| &= \lim_{\rho' \downarrow \rho} \left| \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot \frac{(y-k) r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta, y)}{\rho' - \rho} \right| \\
&\le \lim_{\rho' \downarrow \rho} \int_0^\infty d\Delta \int_k^{k+\Delta} dy \cdot \left| \frac{(y-k) r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta, y)}{\rho' - \rho} \right| \\
&\le \lim_{\rho' \downarrow \rho} \int_0^\infty d\Delta \frac{\Delta}{\rho' - \rho} \int_k^{k+\Delta} dy \cdot |r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta, y)|
\end{aligned}$$

where I've used continuity of the absolute value function and the Minkowski inequality. Define $\xi(\rho,\rho') = \sup_{\epsilon\in E}\Delta(\rho,\rho',\epsilon)$. The strategy will be show that $\lim_{\rho'\downarrow\rho}\xi(\rho,\rho') = 0$, and then since $r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta, y) = 0$ for any $\Delta > \xi(\rho,\rho')$ and all y (since the marginal density $f_{\Delta(\rho,\rho')}(\Delta)$ would be zero for such Δ). With $\xi(\rho,\rho')$ so-defined:

$$\begin{aligned}
\text{RHS of above} &\leq \lim_{\rho'\downarrow\rho}\int_0^{\xi(\rho,\rho')} d\Delta \frac{\xi(\rho,\rho')}{\rho'-\rho}\int_k^{k+\xi(\rho,\rho')} dy\cdot\left|r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,y)\right| \\
&= \lim_{\rho'\downarrow\rho}\frac{\xi(\rho,\rho')}{\rho'-\rho}\cdot\lim_{\rho'\downarrow\rho}\int_0^{\xi(\rho,\rho')} d\Delta\int_0^{\xi(\rho,\rho')} dy\cdot\left|r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,k+y)\right| \quad (23)
\end{aligned}$$

where in the second step I have assumed that each limit exists (this will be demonstrated below). Let us first consider the inner integral of the above: $\int_k^{k+\xi(\rho,\rho')} dy\cdot\left|r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,y)\right|$, for any Δ. Supposing that $\lim_{\rho'\downarrow\rho}\xi(\rho,\rho') = 0$, it follows that this inner integral evaluates to zero, by the Leibniz rule and using that $r_{\Delta_i(\rho,\rho'),Y_i(\rho)}(\Delta,k) = 0$. Thus the entire second limit is equal to zero.

Now I prove that $\lim_{\rho'\downarrow\rho}\xi(\rho,\rho') = 0$ and that $\lim_{\rho'\downarrow\rho}\frac{\xi(\rho,\rho')}{\rho'-\rho}$ exists. First, note that continuous differentiability of $Y(\rho,\epsilon_i)$ implies $Y_i(\rho)$ is continuous for each i so $\lim_{\rho'\downarrow\rho}\Delta_i(\rho,\rho') = 0$ point-wise in ϵ. We seek to turn this point-wise convergence into uniform convergence over ϵ, i.e. that $\lim_{\rho'\downarrow\rho}\sup_{\epsilon\in E}\Delta(\rho,\rho',\epsilon) = \sup_{\epsilon\in E}\lim_{\rho'\downarrow\rho}\Delta(\rho,\rho',\epsilon) = \sup_{\epsilon\in E} 0 = 0$. The strategy will be to use equicontinuity of the sequence and compactness of E. Consider any such sequence $\rho_n \xrightarrow{n} \rho$ from above, and let $f_n(\epsilon) := Y(\rho,\epsilon) - Y(\rho_n,\epsilon)$ and $f(\epsilon) = \lim_{n\to\infty} f_n(\epsilon) = 0$. Equicontinuity of the sequence $f_n(\epsilon)$ says that for any $\epsilon,\epsilon'\in E$ and $e > 0$, there exists a $\delta > 0$ such that $||\epsilon-\epsilon'|| < \delta \implies |f_n(\epsilon) - f_n(\epsilon')| < e$.

This follows from continuous differentiability of $Y(\rho,\epsilon)$. Let $M = \sup_{\rho\in[\rho_0,\rho_1],\epsilon\in E}|\nabla_{\rho,\epsilon}Y(\rho,\epsilon)|$. M exists and is finite given continuity of the gradient and compactness of $[\rho_0,\rho_1]\times E$. Then, for any two points $\epsilon,\epsilon'\in E$ and any $\rho\in[\rho_0,\rho_1]$:

$$|Y(\rho,\epsilon) - Y(\rho,\epsilon')| = \left|\int_{\epsilon'}^{\epsilon}\nabla_\epsilon Y(\rho,\epsilon)\cdot\mathbf{d}\epsilon\right| \leq \int_{\epsilon'}^{\epsilon}|\nabla_\epsilon Y(\rho,\epsilon)\cdot\mathbf{d}\epsilon| \leq M\int_{\epsilon'}^{\epsilon}||\mathbf{d}\epsilon|| \leq M||\epsilon-\epsilon'||$$

where $\mathbf{d}\epsilon$ is any path from ϵ to ϵ' and I have used the definition of M and Cauchy-Schwarz in the second inequality. The existence of a uniform Lipschitz constant M for $Y(\rho,\epsilon)$ implies a uniform equicontinuity of $Y(\rho,\epsilon)$ of the form that for any $e > 0$ and $\epsilon,\epsilon'\in E$, there exists a $\delta > 0$ such that $||\epsilon-\epsilon'|| < \delta \implies \sup_{\rho\in[\rho_0,\rho_1]}|Y(\rho,\epsilon) - Y(\rho,\epsilon')| < e/2$, since we can simply take $\delta = e/(2M)$. This in turn implies that whenever $||\epsilon-\epsilon'|| < \delta$:

$$\begin{aligned}
|Y(\rho,\epsilon) - Y(\rho_n,\epsilon) - \{Y(\rho,\epsilon') - Y(\rho_n,\epsilon')\}| &= |Y(\rho,\epsilon) - Y(\rho,\epsilon') - \{Y(\rho_n,\epsilon) - Y(\rho_n,\epsilon')\}| \\
&\leq |Y(\rho,\epsilon) - Y(\rho,\epsilon')| + |Y(\rho_n,\epsilon) - Y(\rho_n,\epsilon')| \leq e,
\end{aligned}$$

our desired result. Together with compactness of E, equicontinuity implies that $\lim_{n\to\infty}\sup_{\epsilon\in E} f_n(\epsilon) = \sup_{\epsilon\in E}\lim_{n\to\infty} f_n(\epsilon) = 0$.

We apply an analogous argument for $\lim_{\rho'\downarrow\rho}\frac{\xi(\rho,\rho')}{\rho'-\rho}$, where now $f_n(\epsilon) = \frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho}$. For this case it's easier to work directly with the function $\frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho}$, showing that it is Lipschitz in deviations of ϵ uniformly over $n \in \mathbb{N}, \epsilon \in E$.

$$\begin{aligned}
\left|\frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho} - \frac{Y(\rho,\epsilon')-Y(\rho_n,\epsilon')}{\rho_n-\rho}\right| &= \frac{1}{\rho_n-\rho}\left|\int_{\epsilon'}^{\epsilon} \nabla_\epsilon Y(\rho,\epsilon)\cdot \mathbf{d}\epsilon - \int_{\epsilon'}^{\epsilon} \nabla_\epsilon Y(\rho_n,\epsilon)\cdot \mathbf{d}\epsilon\right| \\
&\leq \frac{1}{\rho_n-\rho}\left(\left|\int_{\epsilon'}^{\epsilon} \nabla_\epsilon Y(\rho,\epsilon)\cdot \mathbf{d}\epsilon\right| + \left|\int_{\epsilon'}^{\epsilon} \nabla_\epsilon Y(\rho_n,\epsilon)\cdot \mathbf{d}\epsilon\right|\right) \\
&\leq \frac{2M}{\rho_n-\rho}\int_{\epsilon'}^{\epsilon} ||\mathbf{d}\epsilon|| \leq \frac{2M}{\rho_n-\rho}||\epsilon-\epsilon'||
\end{aligned}$$

This implies equicontinuity of $\frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho}$with the choice $\delta = e(\rho_n-\rho)/(2M)$. As before, equicontinuity and compactness of E allow us to interchange the limit and the supremum, and thus:

$$\begin{aligned}
\lim_{n\to\infty}\frac{\xi(\rho,\rho_n)}{\rho_n-\rho} = \lim_{n\to\infty}\frac{\sup_{\epsilon\in E}\{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)\}}{\rho_n-\rho} &= \lim_{n\to\infty}\sup_{\epsilon\in E}\frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho} \\
= \sup_{\epsilon\in E}\lim_{n\to\infty}\frac{Y(\rho,\epsilon)-Y(\rho_n,\epsilon)}{\rho_n-\rho} &= \sup_{\epsilon\in E}\frac{\partial Y(\rho,\epsilon)}{\partial\rho} := M' < \infty
\end{aligned}$$

where finiteness of M' follows from it being defined as the supremum of a continuous function over a compact set. This establishes that the first limit in Eq. (23) exists and is finite, completing the proof that it evaluates to zero.

Given that the second term in Eq. (22) is zero, we can simplify the remaining term as:

$$\begin{aligned}
\lim_{\rho'\downarrow\rho}\frac{P(Y_i(\rho)\leq k\leq Y_i(\rho'))-p(k)}{\rho'-\rho} &= \lim_{\rho'\downarrow\rho}\frac{1}{\rho'-\rho}\int_0^\infty f_{\Delta(\rho,\rho'),Y(\rho)}(\Delta,k)\Delta d\Delta \\
&= f_\rho(k)\lim_{\rho'\downarrow\rho}\frac{1}{\rho'-\rho}P(\Delta_i(\rho,\rho')\geq 0|Y_i(\rho)=k) \\
&\qquad \cdot\mathbb{E}\left[\Delta_i(\rho,\rho')|Y_i(\rho)=k,\Delta_i(\rho,\rho')\geq 0\right] \\
&= f_\rho(k)(k)\lim_{\rho'\downarrow\rho}\frac{1}{\rho'-\rho}\mathbb{E}\left[\Delta_i(\rho,\rho')|Y_i(\rho)=k,\Delta_i(\rho,\rho')\right] \\
&= f_\rho(k)(k)\mathbb{E}\left[\lim_{\rho'\downarrow\rho}\frac{\Delta_i(\rho,\rho')}{\rho'-\rho}\right| Y_i(\rho)=k\Bigg] \\
&= f_\rho(k)\mathbb{E}\left[-\frac{Y_i(\rho)}{d\rho}\right| Y_i(\rho)=k\Bigg]
\end{aligned}$$

where I have used Lemma POS and then finally the dominated convergence theorem. To see that we may use the latter, note that $\frac{dY_i(\rho)}{d\rho} = \frac{\partial Y(\rho,\epsilon_i)}{\partial\rho} < M$ uniformly over all $\epsilon_i \in E$, and $\mathbb{E}\left[M|\, Y_i(\rho)=k\right] = M < \infty$.

H.4 Proof of Appendix D Proposition 3

Note: this proof follows the notation of Y_i from Appendix A, rather than h_{1it} from Appendix D and the main text. Begin with the following observations:

- $(Y < k) \implies (Y_0 = Y)$ and $(Y > k) \implies (Y_1 = Y)$ both follow from convexity of preferences, and linearity of the cost functions B_1 and B_0. From these two it also follows that $(Y_1 \leq k \leq Y_0) \implies (Y = k)$. See proof of Theorem 1, which treats this case.
- For firm-choosers: $(Y_0 < k) \implies (Y = Y_0)$, since the cost function B_0 coincides with B_k for $y \leq k$, and is higher otherwise. Similarly $(Y_1 > k) \implies (Y = Y_1)$. Together these also imply that $(Y = k) \implies (Y_1 \leq k \leq Y_0)$.
- By analagous logic, for worker-choosers: $(Y_0 \geq k) \implies (Y = Y_1)$, and $(Y_1 \leq k) \implies (Y = Y_0)$ using that their utility functions are strictly increasing in c. Together these also imply that $Y_1 \leq k \leq Y_0$ can only occur if $Y_0 = Y_1 = k$.

Now consider the claims of the Proposition:

- $P(Y_{it} = k \text{ and } K^*_{it} = 0) = P(Y_{1it} \leq 40 \leq Y_{0it} \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 0)$
- $\lim_{y \uparrow 40} f(y) = P(W_{it} = 0) \lim_{y \uparrow 40} f_{0|W=0}(y)$
- $\lim_{y \downarrow 40} f(y) = P(W_{it} = 0) \lim_{y \downarrow 40} f_{1|W=0}(y)$

First claim:

$$\begin{aligned}
P(Y_{it} = k \text{ and } K^*_{it} = 0) &= P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 0) + P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 1) \\
&= P(Y_{1it} \leq 40 \leq Y_{0it} \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 0) + 0
\end{aligned}$$

where for the first term I've used that when $W_{it} = 0$, $(Y_{it} = k) \iff (Y_{1it} \leq 40 \leq Y_{0it})$ following Theorem 1. For the second, I've used that by the absolute continuity assumption: $P(Y_{0it} = k \text{ or } Y_{1it} = k | K^*_{it} = 0) = 0$, so:

$$\begin{aligned}
P(Y_{it} = k \text{ and } K^*_{it} = 0) &= P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 1 \text{ and } Y_{0it} < k) \\
&\qquad + P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 1 \text{ and } Y_{0it} > k) \\
&= P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 1 \text{ and } Y_{0it} < k \text{ and } Y_{1it} = k) \\
&\quad + P(Y_{it} = k \text{ and } K^*_{it} = 0 \text{ and } W_{it} = 1 \text{ and } Y_{0it} > k \text{ and } Y_{1it} = k) \\
&= 0 + 0 = 0
\end{aligned}$$

where I've used that $W_{it} = 1$ and $Y_{0it} < k$ and implies that $Y_{it} = Y_{0it}$ if $Y_{1it} < k$, and $Y_{it} \in \{Y_{0it}, Y_{1it}\}$ if $Y_{1it} > k$ to eliminate the first term. The second term uses that $Y_1 \leq k \leq Y_0$ can only occur when $Y_0 = Y_1 = k$.

Second claim:

$$\begin{aligned}
\lim_{y \uparrow k} f(y) &= \lim_{y \uparrow k} \frac{d}{dy} P(Y_{it} \leq y) \\
&= \lim_{y \uparrow k} \frac{d}{dy} P(Y_{it} \leq y \text{ and } W_{it} = 0) + \lim_{y \uparrow k} \frac{d}{dy} P(Y_{it} \leq y \text{ and } W_{it} = 1)
\end{aligned}$$

The first term is equal to $P(W_{it} = 0)\lim_{y\uparrow k} f_{0|W=0}(y)$, and I now show that the second is equal to zero:

$$
\begin{aligned}
\lim_{y\uparrow k} \frac{d}{dy} & P(Y_{it} \leq y \text{ and } W_{it} = 1) \\
&= \lim_{y\uparrow k} \frac{d}{dy} P(Y_{0it} \leq y \text{ and } Y_{it} = Y_{0it} \text{ and } W_{it} = 1) \\
&= \lim_{y\uparrow k} \frac{d}{dy} P(Y_{0it} \leq y \text{ and } \{u(B_0(Y_{0it}), Y_{0it}) \geq u_{it}(B_1(y), y) \text{ for all } y > k\} \text{ and } W_{it} = 1)
\end{aligned}
$$

For $it's$ utility under B_k at Y_{0it} to be greater than that attainable at any $y > k$, the indifference curve IC_{0it} passing through Y_{0it} must lie above $B_{1it}(y) = w_{it}y + \frac{w_{it}}{2}(y-k)$ for all $y > k$. Using that IC_{0it} passes through the point $(w_{it}Y_{0it}, Y_{0it})$ with derivative w_{it} there (by the first-order condition for an optimum), we may write it as

$$
\begin{aligned}
IC_{0it}(y) = w_{it}Y_{0it} + \int_{Y_{0it}}^{y} IC'_{0it}(y')dy' &= w_{it}Y_{0it} + \int_{Y_{0it}}^{y} \left\{ w_{it} + \int_{Y_{0it}}^{y'} IC''_{0it}(y'')dy'' \right\} dy' \\
\leq w_{it}y + \int_{Y_{0it}}^{y} M(y' - Y_{0it})dy &= w_{it}y + \frac{1}{2}(y - Y_{0it})^2 M_{it}
\end{aligned}
$$

using that IC_{0it} is twice differentiable. Now $IC_{0it}(y) \geq B_{1it}(y)$ for $y > k$ implies that

$$\frac{w_{it}}{M_{it}}(y-k) \leq (y - Y_{0it})^2$$

Taking for example $y = 80 - Y_{0it}$, such that $y - k = y - Y_{0it}$, we have that $Y_{0it} \leq k - \frac{w_{it}}{M_{it}}$. Thus:

$$
\begin{aligned}
\lim_{y\uparrow k} \frac{d}{dy} & P(Y_{it} \leq y \text{ and } Y_{it} > Y_{0it} \text{ and } W_{it} = 1) \\
&\leq \lim_{y\uparrow k} \lim_{\delta\downarrow 0} \frac{1}{\delta} P(Y_{0it} \in (y-\delta, y] \text{ and } Y_{0it} \leq k - \frac{w_{it}}{M_{it}} \text{ and } W_{it} = 1) \\
&\leq \lim_{y\uparrow k} \lim_{\delta\downarrow 0} \frac{1}{\delta} P(Y_{0it} \in (y-\delta, y] \text{ and } \frac{w_{it}}{M_{it}} \leq k - y + \delta \text{ and } W_{it} = 1) \\
&\leq \lim_{y\uparrow k} \lim_{\delta\downarrow 0} \frac{1}{\delta} P(\frac{w_{it}}{M_{it}} \leq k - y + \delta \text{ and } W_{it} = 1) \\
&\leq \lim_{\delta\downarrow 0} \frac{1}{\delta} P\left(\frac{w_{it}}{M_{it}} \leq \delta \text{ and } W_{it} = 1\right) \\
&= f_{w/m|W=1}(0) = 0
\end{aligned}
$$

where we may interchange the limits given that $\frac{w_{it}}{M_{it}}$ conditional on $W_{it} = 1$ admits a density $f_{w/m|W=1}$ that is bounded in a neighborhood around 0. This, and that $f_{w/m|W=1}(0) = 0$ follows from the assumption that the distribution of M_{it}/w_{it} is bounded.

We have now proved the second claim, that $\lim_{y\uparrow k} f(y) = P(W_{it} = 0)\lim_{y\uparrow k} f_{0|W=0}(y)$.

<u>Third claim:</u> Analogous logic to the second claim, using the bounded 2^{nd} derivative of IC_{1it}.

H.5 Proof of Appendix D Theorem 1*

Note: this proof follows the notation of Y_i from Appendix A, rather than h_{1it} from Appendix D and the main text. Let $T_i = 1$ be a shorthand for firm-choosers who are not counterfactual bunchers, i.e. the event $K_{it}^* = 0$ and $W_{it} = 0$.

By Theorem 1 of Dümbgen et al., 2017: for $d \in \{0, 1\}$ and any t, bi-log concavity implies that:

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

Defining $u = F_{0|T=1}(k+t)$, we can use the substitution $t = Q_{0|T=1}(u) - k$ to translate the above into bounds on the conditional quantile function of Y_{0i}, evaluated at u:

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

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

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

By RANK, we have that $Y_i = k \iff F_{0|T=1}(Y_{0i}) \in [F_{0|T=1}(k), F_{0|T=1}(k) + \mathcal{B}^*] \iff F_{1|T=1}(Y_{1i}) \in [F_{1|T=1}(k) - \mathcal{B}^*, F_{1|T=1}(k)]$ where $\mathcal{B}^* := P(Y_i = k|T = 1)$, and thus:

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

A lower bound for $E[Y_{0i} - Y_{1i}|Y_i = k, T_i = 0]$ is thus:

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

where as in Theorem 1: $g(a, b, x) = \frac{a}{bx}(a + x)\ln\left(1 + \frac{x}{a}\right) - \frac{a}{b}$ and $h(a, b, x) = g(1 - a, b, x)$. Similarly, an upper bound is:

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

where again $\tilde{g}(a, b, x) = -g(1 - a, b, -x)$ and $\tilde{h}(a, b, x) = -g(a, b, -x)$. We have then that $E[Y_{0i} - Y_{1i}|Y_i = k, T_i = 0] \in [\Delta_k^L, \Delta_k^U]$, where:

$$\begin{aligned}\Delta_k^L &= g(F_{0|T=1}(k), f_{0|T=1}(k), \mathcal{B}^*) + g(1 - F_{1|T=1}(k), f_{1|T=1}(k), \mathcal{B}^*) \\ &= g\left(P(Y_{0i} \leq k \text{ and } T_i = 1), P(T_i = 1)f_{0|T=1}(k), P(Y_i = k \text{ and } T_i = 1)\right) \\ &\qquad + g\left(P(Y_{1i} > k \text{ and } T_i = 1), P(T_i = 1)f_{1|T=1}(k), P(Y_i = k \text{ and } T_i = 1)\right)\end{aligned}$$

$$
\begin{aligned}
\Delta_k^U &= -g(1-F_{0|T=1}(k), f_{0|T=1}(k), -\mathcal{B}^*) - g(F_{1|T=1}(k), f_{1|T=1}(k), -\mathcal{B}^*) \\
&= -g\left(P(Y_{0i} > k \text{ and } T_i = 1), P(T_i = 1) f_{0|T=1}(k), -P(Y_i = k \text{ and } T_i = 1)\right) \\
&\qquad\qquad - g\left(P(Y_{1i} \leq k \text{ and } T_i = 1), P(T_i = 1) f_{1|T=1}(k), -P(Y_i = k \text{ and } T_i = 1)\right)
\end{aligned}
$$

where I've used that the function $g(a, b, x)$ is homogeneous of degree zero and multiplied each argument by $P(T_i = 1)$. The bounds are sharp as CHOICE, CONVEX and RANK imply no further restrictions on the marginal potential outcome distributions.

Next, note that:

$$
\begin{aligned}
\lim_{y \uparrow k} f(y) &= \lim_{y \uparrow k} \frac{d}{dy} P(Y_{0i} \leq y \text{ and } W_i = 0) = \lim_{y \uparrow k} \frac{d}{dy} P(Y_{0i} \leq y \text{ and } W_i = 0 \text{ and } K_i^* = 0) \\
&= P(T_i = 1) \cdot \lim_{y \uparrow k} \frac{d}{dy} P(Y_{0i} \leq y | T_i = 1) = P(T_i = 1) \cdot f_{0|T=1}(k)
\end{aligned}
$$

$$
\begin{aligned}
\lim_{y \downarrow k} f(y) &= -\lim_{y \downarrow k} \frac{d}{dy} P(Y_{1i} \geq y \text{ and } W_i = 0) = -\lim_{y \downarrow k} \frac{d}{dy} P(Y_{1i} \geq y \text{ and } W_i = 0 \text{ and } K_i^* = 0) \\
&= P(T_i = 1) \cdot -\lim_{y \downarrow k} \frac{d}{dy} P(Y_{1i} \geq y | T_i = 1) = P(T_i = 1) \cdot f_{1|T=1}(k)
\end{aligned}
$$

$$
\mathcal{B} - p = P(Y_i = k \text{ and } K_i^* = 0) = P(Y_i = k \text{ and } K_i^* = 0 \text{ and } W_i = 0) = P(Y_i = k \text{ and } T_i = 1)
$$

As shown by Dümbgen et al., 2017, BLC implies the existence of a continuous density function, which assures that these density limits exist and are equal to the corresponding potential outcome densities above. Thus, the quantities $P(Y_i = k \text{ and } T_i = 1)$, $P(T_i = 1) \cdot f_{0|T=1}(k)$ and $P(T_i = 1) \cdot f_{1|T=1}(k)$ are all point-identified from the data.

Now we turn to the CDF arguments of Δ_k^L and Δ_k^U. Note that the desired quantities can be written

- $P(Y_{0i} \leq k \text{ and } T_i = 1) = P(Y_{0i} < k \text{ and } T_i = 1) = P(Y_{0i} < k \text{ and } W_i = 0)$
- $P(Y_{1i} > k \text{ and } T_i = 1) = P(Y_{1i} > k \text{ and } W_i = 0)$
- $P(Y_{0i} > k \text{ and } T_i = 1) = P(Y_{0i} > k \text{ and } W_i = 0)$
- $P(Y_{1i} \leq k \text{ and } T_i = 1) = P(Y_{1i} < k \text{ and } T_i = 1) = P(Y_{1i} < k \text{ and } W_i = 0)$

Let

$$
A := P(Y_{0i} < k \text{ and } Y_i = Y_{0i} \text{ and } W_i = 1) \quad \text{and} \quad B := P(Y_{1i} > k \text{ and } Y_i = Y_{1i} \text{ and } W_i = 1)
$$

The desired quantities are related to observables via A and B:

- $P(Y_i < k) = P(Y_{0i} < k \text{ and } W_i = 0) + A$
- $P(Y_i > k) = P(Y_{1i} > k \text{ and } W_i = 0) + B$
- $P(Y_i \leq k) - p = P(Y_i \leq k \text{ and } K_i^* = 0) = P(Y_i \leq k \text{ and } T_i = 1) + A = P(Y_{1i} \leq k \text{ and } W_i = 0) + A$

- $P(Y_i \geq k) - p = P(Y_i \geq k \text{ and } K_i^* = 0) = P(Y_i \geq k \text{ and } T_i = 1) + B = P(Y_{0i} > k \text{ and } W_i = 0) + B$

The four CDF arguments appearing in Δ_k^L and Δ^U are thus identified up to the correction terms A and B. A simple sufficient condition for $A = B = 0$ is that there are no worker-choosers.

H.6 Proof of Appendix F.1 Proposition 4

The first order conditions with respect to z and h are:

$$\lambda F_L(L, K)e(h) = \phi + \frac{\beta_Y(z, h) + 1}{\beta_Y(z, h)} z$$

and

$$\lambda F_L(L, K)e(h)(\eta(h)/\beta_h(z, h) + 1) = z + \phi$$

where $L = N(z, h)e(h)$, $\eta(h) := e'(h)h/e(h)$, $\beta_h(z, h) := N_h(z, h)h/N(z, h)$ and $\beta_z(z, h) := N_z(z, h)Y/N(z, h)$ are elasticity functions and λ is a Lagrange multiplier. I have assumed that the functions $|\beta_h|$, β_h, and η are strictly positive and finite globally. Combining the two equations, we have that an interior solution must satisfy either: $z = \frac{\phi \frac{\eta}{\beta_h}}{1 - \frac{\beta_z+1}{\beta_z} \frac{\beta_h+\eta}{\beta_h}}$ (Case 1), or that the denominator of the above is zero: $\frac{\beta_h}{\beta_h+\eta} = \frac{\beta_z+1}{\beta_z}$ (Case 2), where the dependence of β_z and β_h has been left implicit. Defining $\beta(z, h) = |\beta_h(z, h)|/(\beta_z(z, h) + 1)$, we can rewrite the condition for Case 2 as $\beta(z, h) = \eta(h)$.

With $\phi = 0$, we must be in Case 2 for any $z > 0$ to have positive profits, and not that positivity of z requires $\beta < \eta$ in case one. On the other hand if $\phi > 0$ we cannot have Case 1 provided that $\eta/\beta_h > 0$. Now specialize to the conditions set out in the Proposition: that $F_L = 1$, $\lambda = 1$ (profit maximization), and β_h, β_z and η are all constants. Then $z = \frac{\phi \frac{\eta}{\beta_h}}{1 - \frac{\beta_z+1}{\beta_z} \frac{\beta_h+\eta}{\beta_h}} = \phi \cdot \frac{\beta_z}{\beta_z+1}$ and the first order condition for hours becomes

$$e(h) = \phi + \phi \frac{\eta}{\beta - \eta}$$

which simplifies to $h = \left[\frac{\phi}{e_0} \cdot \frac{\beta}{\beta - \eta}\right]^{1/\eta}$.

H.7 Proof of Appendix Proposition 5

By constant treatment effects, $f_1^G(y) = f_0^G(y + \delta)$ and note that both $f_0^G(k)$ and $f_1^G(k)$ are identified from the data. These can be transformed into densities for Y_{0i} and Y_{1i} via $f_d(y) = G'(y) f_d^G(G(y))$ for $d \in \{0, 1\}$. With $f_0(y)$ linear on the interval $[k, k + \Delta]$, the integral $\int_k^{k+\Delta} f_0(y)dy$ evaluates to $\mathcal{B} = \frac{\Delta}{2}(f_0(k) + f_0(k + \Delta))$. Although $f_0(k) = \lim_{y \uparrow k} f(y)$ by CONT, $f_0(k + \Delta)$ is not immediately observable. However:

$$f_0(k + \Delta) = f_0\left(G^{-1}(G(k) + \delta)\right) = G'(k + \Delta) f_0^G(G(k) + \delta)$$

and furthermore by constant treatment effects:

$$f_0^G(G(k)+\delta) = f_1^G(G(k)) = (G'(k))^{-1} f_1(k) = (G'(k))^{-1} \lim_{y \downarrow k} f(y)$$

Combining these equations, we have the result.

H.8 Proof of Appendix Proposition 6

We seek a Δ such that for some θ_0:

$$\mathcal{B} = \int_{\tilde{k}}^{k+\Delta} g(y;\theta_0)dy \tag{24}$$

and

$$f(y) = \begin{cases} g(y;\theta_0) & y < k \\ g(y+\Delta;\theta_0) & y > k \end{cases} \tag{25}$$

and

$$g(y;\theta_0) > 0 \text{ for all } y \in [k, k+\Delta] \tag{26}$$

Recall from Equation (17) that $\Delta = G^{-1}(G(k)+\delta) - k$ and hence $\delta = G(k+\Delta) - G(k)$. Thus if we find a unique Δ satisfying the two equations, we have found a unique value of δ: the true value of the homogenous effect δ^G.

Suppose we have two candidate values $\Delta' > \Delta$. For them to both satisfy (24), we would need $\Delta' = \Delta(\theta')$ and $\Delta = \Delta(\theta)$, where $\theta, \theta' \in \Theta$ and $\Delta(\theta_0)$ is the unique Δ satsifying Eq. (24) for a given θ_0, which is unique for each permissible value θ_0 by the positivity condition (26). To satisfy (25), we would also need

$$g(y;\theta) = \begin{cases} f(y) & y < k \\ f(y-\Delta(\theta)) & y > k+\Delta(\theta) \end{cases} \qquad g(y;\theta') = \begin{cases} f(y) & y < k \\ f(y-\Delta(\theta')) & y > k+\Delta(\theta') \end{cases} \tag{27}$$

Since $g(y;\theta)$ is a real analytic function for any $\theta \in \Theta$, the function $h_{\theta\theta'}(y) := g(y;\theta) - g(y;\theta')$ is real analytic. An implication of this is that if $h_{\theta\theta'}(y)$ vanishes on the interval $[0, \tilde{k}]$, as it must by Equation (27), it must vanish everywhere on $\mathbb{R}$. Thus for any $y > k + \Delta(\theta)$:

$$g(y+\Delta(\theta')-\Delta(\theta);\theta) = g(y+\Delta(\theta')-\Delta(\theta);\theta') = g(y;\theta)$$

So $g(y;\theta)$ is periodic with period $\Delta(\theta') - \Delta(\theta)$. Since g is non-negative, it cannot integrate to unity globally, and thus cannot be the same function as $f_0(y)$.

H.9 Details of calculations for policy estimates

H.9.1 Ex-post evaluation of time-and-a-half after 40

$$\mathbb{E}[Y_{0i} - Y_i] = (\mathcal{B} - p)\mathbb{E}[Y_{0i} - k | Y_i = k, K_i^* = 0] + p \cdot 0 + P(Y_{1i} > k)\mathbb{E}[Y_{0i} - Y_{1i} | Y_i > k]$$

Consider the first term

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

where $\mathcal{B}^*:=P(Y_i=k|K^*=0)=\frac{\mathcal{B}-p}{1-p}$. Bounds for the rightmost quantity are given by bi-log-concavity of Y_{0i}, just as in Theorem 1. In particular:

$$\begin{aligned}(\mathcal{B}-p)E[Y_{0i}-k|Y_i=k,K_i^*=0]&\geq(1-p)\mathcal{B}^*\cdot\frac{F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)(\mathcal{B}^*)}\int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k)+\mathcal{B}^*}\ln\left(\frac{u}{F_{0|K^*=0}(k)}\right)du\\&=(1-p)\mathcal{B}^*\cdot g(F_{0|K^*=0}(k),f_{0|K^*=0}(k),\mathcal{B}^*)\\&=(\mathcal{B}-p)\cdot g(F_-,f_-,\mathcal{B}-p)\end{aligned}$$

and

$$\begin{aligned}(\mathcal{B}-p)E[Y_{0i}-k|Y_i=k,K_i^*=0]&\leq-(1-p)\mathcal{B}^*\cdot\frac{1-F_{0|K^*=0}(k)}{f_{0|K^*=0}(k)(\mathcal{B}^*)}\int_{F_{0|K^*=0}(k)}^{F_{0|K^*=0}(k)+\mathcal{B}^*}\ln\left(\frac{1-u}{1-F_{0|K^*=0}(k)}\right)du\\&=(1-p)\mathcal{B}^*\cdot g'(F_{0|K^*=0}(k),f_{0|K^*=0}(k),\mathcal{B}^*)\\&=-(\mathcal{B}-p)\cdot g(1-p-F_-,f_+,p-\mathcal{B})\end{aligned}$$

where as before $g(a,b,x)=\frac{a}{bx}(a+x)\ln\left(1+\frac{x}{a}\right)-\frac{a}{b}$ and $g'(a,b,x)=-g(1-a,b,-x)$.

Now consider the second term of $\mathbb{E}[Y_{0i}-Y_i]$: $P(Y_{1i}>k)\mathbb{E}[Y_{0i}-Y_{1i}|Y_i>k]$. Taking as a lower bound an assumption of constant treatment effects in levels: $P(Y_{1i}>k)\mathbb{E}[Y_{0i}-Y_{1i}|Y_i>k]\geq P(Y_{1i}>k)\Delta_k^L$.

For an upper bound, we assume that $\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{\rho}{Y_i(\rho)}\middle|Y_i(\rho')=y,K_i^*=0\right]=\mathcal{E}$ for all ρ, ρ' and y. Consider then the buncher ATE in logs:

$$\begin{aligned}\mathbb{E}\left[\ln Y_{0i}-\ln Y_{1i}|Y_i=k,K_i^*=0\right]&=\mathbb{E}\left[\ln Y_{0i}-\ln Y_{1i}|Y_{0i}\in[k,Q_{0|K^*=0}(F_{1|K^*=0})],K_i^*=0\right]\\&=\int_{\rho_0}^{\rho_1}d\rho\cdot\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{1}{Y_i(\rho)}\middle|Y_{0i}\in[k,k+\Delta_0^*],K_i^*=0\right]\\&=\int_{\rho_0}^{\rho_1}d\ln\rho\cdot\frac{1}{\mathcal{B}^*}\int_k^{k+\Delta_0^*}dy\cdot f_0(y)\cdot\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{\rho}{Y_i(\rho)}\middle|Y_{0i}=y,K_i^*=0\right]\\&=\mathcal{E}\int_{\rho_0}^{\rho_1}d\ln\rho=\mathcal{E}\ln(\rho_1/\rho_0)\end{aligned}$$

with the notation that $\Delta_0^*:=Q_{0|K^*=0}(F_{1|K^*=0})-k$. Moreover:

$$\begin{aligned}\mathbb{E}[Y_{0i}-Y_{1i}|Y_i>k]&=\int_{\rho_0}^{\rho_1}d\rho\cdot\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\middle|Y_{1i}>k,K_i^*=0\right]\\&=P(Y_{1i}>k)^{-1}\int_{\rho_0}^{\rho_1}d\ln\rho\cdot\int_k^\infty y\cdot f_1(y)\cdot\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{\rho}{Y_i(\rho)}\middle|Y_{1i}=y,K_i^*=0\right]dy\\&=\mathcal{E}\cdot\mathbb{E}[Y_{1i}|Y_{1i}>k]\int_{\rho_0}^{\rho_1}d\ln\rho=\mathcal{E}\ln(\rho_1/\rho_0)\cdot\mathbb{E}[Y_{1i}|Y_{1i}>k]\end{aligned}$$

Thus in the isoelastic model

$$E[Y_{0i}-Y_i] = (\mathcal{B}-p)E[Y_{0i}-k|Y_i = k, K_i^* = 0]+\mathbb{E}[Y_{1i}|Y_{1i} > k]\cdot P(Y_{1i} > k)\mathbb{E}\left[\ln Y_{0i} - \ln Y_{1i}|Y_i = k, K_i^* = 0\right]$$

and an upper bound is

$$\delta_k^U \cdot E[Y_i|Y_i > k] - (\mathcal{B} - p) \cdot g(1 - p - F_-, f_+, p - \mathcal{B})$$

where δ_k^U is an upper bound to the buncher ATE in logs $\mathbb{E}\left[\ln Y_{0i} - \ln Y_{1i}|Y_i = k, K_i^* = 0\right]$.

H.9.2 Moving to double time

I make use of the first step deriving the expression for $\partial_{\rho_1} E[Y_i^{[k,\rho_1]}]$ in Theorem 2, namely that:

$$\partial_{\rho_1} E[Y_i^{[k,\rho_1]}] = k\partial_{\rho_1}\mathcal{B}^{[k,\rho_1]} + \partial_{\rho_1}\left\{P(Y_i(\rho_1) > k)\mathbb{E}[Y_i(\rho_1)|Y_i(\rho_1) > k]\right\}$$

Thus:

$$\begin{aligned}
E[Y_i^{[k,\rho_1]}] - E[Y_i^{[k,\bar{\rho}_1]}] &= -\int_{\rho_1}^{\bar{\rho}_1} \partial_\rho E[Y_i^{[k,\rho]}]d\rho = -\int_{\rho_1}^{\bar{\rho}_1}\Big\{k\partial_\rho\mathcal{B}^{[k,\rho]} + \partial_\rho\left\{P(Y_i(\rho) > k)\mathbb{E}[Y_i(\rho)|Y_i(\rho) > k]\right\}\Big\}\,d\rho \\
&= -k(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]}) + P(Y_i(\rho_1) > k)\mathbb{E}[Y_i(\rho_1)|Y_i(\rho_1) > k] - P(Y_i(\bar{\rho}_1) > k)\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] \\
&= -k(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]}) + \{P(Y_i(\rho_1) > k) - P(Y_i(\bar{\rho}_1) > k)\}\cdot\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] \\
&\qquad + P(Y_i(\rho_1) > k)\left(\mathbb{E}[Y_i(\rho_1)|Y_i(\rho_1) > k] - \mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k]\right) \\
&= \left(\mathbb{E}[Y_{1i}|Y_{1i} > k] - k\right)(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]}) + P(Y_{1i} > k)\left(\mathbb{E}[Y_{1i}|Y_{1i} > k] - \mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k]\right) \\
&\leq \left(\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] - k\right)(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]}) + P(Y_{1i} > k)\mathbb{E}[Y_i(\rho_1) - Y_i(\bar{\rho}_1)|Y_{1i} > k] \\
&\leq \left(\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] - k\right)(\mathcal{B}^{[k,\rho_1]} - p) + P(Y_{1i} > k)\mathbb{E}[Y_i(\rho_1) - Y_i(\bar{\rho}_1)|Y_{1i} > k] \\
&\leq \left(\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] - k\right)(\mathcal{B}^{[k,\rho_1]} - p) + P(Y_{1i} > k)\mathbb{E}[Y_{0i} - Y_{1i}|Y_{1i} > k] \\
&\approx \left(\mathbb{E}[Y_{1i}|Y_{1i} > k] - k\right)(\mathcal{B}^{[k,\rho_1]} - p) + P(Y_{1i} > k)\mathbb{E}[Y_{0i} - Y_{1i}|Y_{1i} > k] \\
&\leq \left(\mathbb{E}[Y_{1i}|Y_{1i} > k] - k\right)(\mathcal{B}^{[k,\rho_1]} - p) + P(Y_{1i} > k)E[Y_i|Y_i > k]\cdot\delta_k^U
\end{aligned}$$

In the iso-elastic model, making use instead of the final expression for $\partial_{\rho_1} E[Y_i^{[k,\rho_1]}]$ in Thm. 2:

$$\begin{aligned}
&E[Y_i^{[k,\rho_1]}] - E[Y_i^{[k,\bar{\rho}_1]}] = -\int_{\rho_1}^{\bar{\rho}_1} \partial_\rho E[Y_i^{[k,\rho_1]}]d\rho = \int_{\rho_1}^{\bar{\rho}_1} d\rho \int_k^\infty f_\rho(y)\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\bigg| Y_i(\rho) = y\right]dy \\
&= \int_{\rho_1}^{\bar{\rho}_1} d\ln\rho \int_k^\infty f_\rho(y)y\cdot\mathbb{E}\left[\frac{dY_i(\rho)}{d\rho}\frac{\rho}{Y_i(\rho)}\bigg| Y_i(\rho) = y\right]dy \\
&\geq \mathcal{E}\int_{\rho_1}^{\bar{\rho}_1} d\ln\rho \int_k^\infty f_\rho(y)y\cdot dy = \mathcal{E}\int_{\rho_1}^{\bar{\rho}_1} d\ln\rho\cdot P(Y_i(\rho) > k)\mathbb{E}[Y_i(\rho)|Y_i(\rho) > k] \\
&\geq \mathcal{E}\ln(\bar{\rho}_1/\rho_1)\cdot P(Y_i(\bar{\rho}_1) > k)\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] \\
&= \mathcal{E}\ln(\bar{\rho}_1/\rho_1)\cdot\{P(Y_{1i} > k)\mathbb{E}[Y_{1i}|Y_{1i} > k] + (P(Y_i(\bar{\rho}_1) > k)\mathbb{E}[Y_i(\bar{\rho}_1)|Y_i(\bar{\rho}_1) > k] - P(Y_{1i} > k)\mathbb{E}[Y_{1i}|Y_{1i} > k])\} \\
&= \mathcal{E}\ln(\bar{\rho}_1/\rho_1)\cdot\Big\{P(Y_{1i} > k)\mathbb{E}[Y_{1i}|Y_{1i} > k] - \Big(E[Y_i^{[k,\rho_1]}] - E[Y_i^{[k,\bar{\rho}_1]}]\Big) + k(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]})\Big\}
\end{aligned}$$

where in the fourth step I've used that $Y_i(\rho)$ is decreasing in ρ with probability one, which follows from SEPARABLE and CONVEX. So:

$$\begin{aligned} E[Y_i^{[k,\rho_1]}] - E[Y_i^{[k,\bar{\rho}_1]}] &\geq \frac{\mathcal{E}\ln(\bar{\rho}_1/\rho_1)}{1+\mathcal{E}\ln(\bar{\rho}_1/\rho_1)} \cdot \left\{P(Y_{1i} > k)\mathbb{E}[Y_{1i}|Y_{1i} > k] + k(\mathcal{B}^{[k,\bar{\rho}_1]} - \mathcal{B}^{[k,\rho_1]})\right\} \\ &\geq \frac{\mathcal{E}\ln(\bar{\rho}_1/\rho_1)}{1+\mathcal{E}\ln(\bar{\rho}_1/\rho_1)} \cdot P(Y_{1i} > k)\mathbb{E}[Y_{1i}|Y_{1i} > k] \end{aligned}$$

H.9.3 Effect of a change to the kink point on bunching

Using that $p(k^*) = p$ and $p(k') = 0$:

$$\begin{aligned} \mathcal{B}^{[k',\rho_1]} - \mathcal{B}^{[k^*,\rho_1]} &= \left(\mathcal{B}^{[k',\rho_1]} - p(k')\right) - \left(\mathcal{B}^{[k^*,\rho_1]} - p(k^*)\right) - p = -p + \int_{k^*}^{k'} dk \cdot \partial_k \left(\mathcal{B}^{[k',\rho_1]} - p(k)\right) \\ &= -p + \int_{k^*}^{k'} dk \cdot (f_1(k) - f_0(k)) = -p + F_1(k') - F_1(k^*) - F_0(k') + F_0(k^*) \\ &= P(k^* < Y_{1i} \leq k') - P(k^* < Y_{0i} \leq k') - p \\ &= P(k^* < Y_i \leq k') - P(k^* < Y_{0i} \leq k') - p \end{aligned}$$

if $k' > k^*$.
Similarly, if $k' < k^*$:

$$\mathcal{B}^{[k',\rho_1]} - \mathcal{B}^{[k^*,\rho_1]} = P(k' \leq Y_{0i} < k^*) - P(k' \leq Y_{1i} < k^*) - p = P(k' \leq Y_i < k^*) - P(k' \leq Y_{1i} < k^*) - p$$

The Lemma in the next section gives identified bounds on the potential outcome probability in either case.

H.9.4 Average effect of a change to the kink point on hours

$$\begin{aligned} E[Y_i^{[k',\rho_1]}] - E[Y_i^{[k^*,\rho_1]}] &= \int_{k^*}^{k'} \partial_k E[Y_i^{[k,\rho_1]}] dk = \int_{k^*}^{k'} \left\{\mathcal{B}^{[k,\rho_1]} - p(k)\right\} dk \\ &= k\left(\mathcal{B}^{[k,\rho_1]} - p(k)\right)\Big|_{k^*}^{k'} - \int_{k^*}^{k'} k \cdot \partial_k \left\{\mathcal{B}^{[k,\rho_1]} - p(k)\right\} dk \\ &= k'\mathcal{B}^{[k',\rho_1]} - k^*(\mathcal{B} - p) - \int_{k^*}^{k'} y\left(f_1(y) - f_0(y)\right) dy \\ &= (k' - k^*)\mathcal{B}^{[k',\rho_1]} + k^*\left(\mathcal{B}^{[k',\rho_1]} - \mathcal{B}\right) + pk^* - \int_{k^*}^{k'} y\left(f_1(y) - f_0(y)\right) dy \end{aligned}$$

For $k' > k^*$, this is equal to

$$\begin{aligned} &(k' - k^*)\mathcal{B}^{[k',\rho_1]} + k^*\left(\mathcal{B}^{[k',\rho_1]} - (\mathcal{B} - k)\right) + P(k^* < Y_{0i} \leq k')(\mathbb{E}[Y_{0i}|k^* < Y_{0i} \leq k'] \\ &\qquad\qquad - P(k^* < Y_{1i} \leq k')(\mathbb{E}[Y_{1i}|k^* < Y_{1i} \leq k'] \\ &= (k' - k^*)\mathcal{B}^{[k',\rho_1]} + P(k^* < Y_{0i} \leq k')(\mathbb{E}[Y_{0i}|k^* < Y_{0i} \leq k'] - k^*) - P(k^* < Y_{1i} \leq k')(\mathbb{E}[Y_{1i}|k^* < Y_{1i} \leq k'] - k^*) \\ &= (k' - k^*)\mathcal{B}^{[k',\rho_1]} + P(k^* < Y_{0i} \leq k')(\mathbb{E}[Y_{0i}|k^* < Y_{0i} \leq k'] - k^*) - P(k^* < Y_i \leq k')(\mathbb{E}[Y_i|k^* < Y_i \leq k'] - k^*) \end{aligned}$$

The first term represents the mechanical effect from the bunching mass under k' being transported from k^* to k', and can be bounded given the bounds for $\mathcal{B}^{[k',\rho_1]} - \mathcal{B}^{[k^*,\rho_1]}$ in the last section. The last term is point identified from the data, while the middle term can be bounded using bi-log concavity of Y_{0i} conditional on $K^* = 0$. Similarly, when $k' < k^*$, the effect on hours is:

$$\begin{aligned}&(k'-k^*)\mathcal{B}^{[k',\rho_1]} + P(k' \le Y_{0i} < k^*)(k^* - \mathbb{E}[Y_{0i}|k' \le Y_{0i} < k^*]) - P(k' \le Y_{1i} < k^*)(k^* - \mathbb{E}[Y_{1i}|k' \le Y_{1i} < k^*]) \\ &= (k'-k^*)\mathcal{B}^{[k',\rho_1]} + P(k' \le Y_i < k^*)(k^* - \mathbb{E}[Y_i|k' \le Y_i < k^*]) - P(k' \le Y_{1i} < k^*)(k^* - \mathbb{E}[Y_{1i}|k' \le Y_{1i} < k^*])\end{aligned}$$

with the middle term point identified from the data and last term bounded by bi-log concavity of Y_{1i} conditional on $K^* = 0$. The analytic bounds implied by BLC in each case are given by the Lemma below.

Lemma. *Suppose Y_i is a bi-log concave random variable with CDF $F(y)$. Let $F_0 := F(y_0)$ and $f_0 = f(y_0)$ be the CDF and density, respectively, evaluated at a fixed y_0. Then, for any $y' > y_0$:*

$$A \le P(y_0 \le Y_i \le y')\left(\mathbb{E}[Y_i|y_0 \le Y_i \le y'] - y_0\right) \le B$$

and for any $y' < y_0$:

$$B \le P(y' \le Y_i \le y_0)\left(y_0 - \mathbb{E}[Y_i|y' \le Y_i \le y_0]\right) \le A$$

where $A = g(F_0, f_0, F_L(y'))$ and $B = g(1-F_0, f_0, 1-F_U(y'))$, with

$$F_L(y') = 1 - (1-F_0)e^{-\frac{f_0}{1-F_0}(y-y_0)}, \qquad\qquad F_U(y') = F_0 e^{\frac{f_0}{F_0}(y'-y_0)}$$

and

$$g(a,b,c) = \begin{cases} \frac{ac}{b}\left(\ln\left(\frac{c}{a}\right) - 1\right) + \frac{a^2}{b} & \text{if } c > 0 \\ \frac{a^2}{b} & \text{if } c \le 0 \end{cases}$$

In either of the two cases $\max\{0, F_L(y')\} \le F(y') \le \min\{1, F_U(y')\}$.

Proof. As shown by Dümbgen et al., 2017, bi-log concavity of Y_i implies not only that $f(y)$ exists, but that it is strictly positive, and we may then define a quantile function $Q = F^{-1}$ such that $Q(F(y)) = y$ and $y = Q(F(y))$. Theorem 1 of Dümbgen et al., 2017 also shows that for any y':

$$\underbrace{1 - (1-F_0)e^{-\frac{f_0}{1-F_0}(y-y_0)}}_{:=F_L(y')} \le F(y') \le \underbrace{F_0 e^{\frac{f_0}{F_0}(y'-y_0)}}_{:=F_U(y')}$$

We can re-express this as bounds on the quantile function evaluated at any $u' \in [0,1]$:

$$\underbrace{y_0 + \frac{F_0}{f_0}\ln\left(\frac{u}{F_0}\right)}_{Q_L(u')} \le Q(u') \le \underbrace{y_0 - \frac{1-F_0}{f_0}\ln\left(\frac{1-u}{1-F_0}\right)}_{Q_U(u')}$$

Write the quantity of interest as:

$$P(y_0 \le Y_i \le y')\left(\mathbb{E}[Y_i|y_0 \le Y_i \le y'] - y_0\right) = \int_{y_0}^{y'}(y-y_0)f(y)dy = \int_{F_0}^{F(y')}(Q(u)-y_0)du$$

Given that $Q(u) \geq y_0$, the integral is increasing in $F(y')$. Thus an upper bound is:

$$
\begin{aligned}
P(y_0 \leq Y_i \leq y')\left(\mathbb{E}[Y_i|y_0 \leq Y_i \leq y'] - y_0\right) &\leq \int_{F_0}^{F_U(y')} (Q_U(u) - y_0)du \\
&= -\frac{1-F_0}{f_0}\int_{F_0}^{F_U(y')} \ln\left(\frac{1-u}{1-F_0}\right)du \\
&= \frac{(1-F_0)^2}{f_0}\int_1^{\frac{1-F_U(y')}{1-F_0}} \ln(v)\,dv \\
&= \frac{(1-F_0)(1-F_U(y'))}{f_0}\left(\ln\left(\frac{1-F_U(y')}{1-F_0}\right) - 1\right) + \frac{(1-F_0)^2}{f_0}
\end{aligned}
$$

where we've made the substitution $v = \frac{1-u}{1-F_0}$ and used that $\int \ln(v)dv = v(\ln(v) - 1))$. Inspection of the formulas for F_U and F_L reveal that $F_U \in (0, \infty)$ and $F_L \in (-\infty, 1)$. In the event that $F_U(y') \geq 1$, the above expression is undefined but we can replace $F_U(y')$ with one and still obtain valid bounds:

$$
P(y_0 \leq Y_i \leq y')\left(\mathbb{E}[Y_i|y_0 \leq Y_i \leq y'] - y_0\right) \leq -\frac{(1-F_0)^2}{f_0}\int_0^1 \ln(v)\,dv = \frac{(1-F_0)^2}{f_0}
$$

where we've used that $\int_0^1 \ln(v)dv = -1$.

Similarly, a lower bound is:

$$
\begin{aligned}
P(y_0 \leq Y_i \leq y')\left(\mathbb{E}[Y_i|y_0 \leq Y_i \leq y'] - y_0\right) &\geq \int_{F_0}^{F_L(y')} (Q_L(u) - y_0)du = \frac{F_0}{f_0}\int_{F_0}^{F_L(y')} \ln\left(\frac{u}{F_0}\right)du \\
&= \frac{F_0^2}{f_0}\int_1^{F_L(y')/F_0} \ln(v)\,du \\
&= \frac{F_0 F_L(y')}{f_0}\left(\ln\left(\frac{F_L(y')}{F_0}\right) - 1\right) + \frac{F_0^2}{f_0}
\end{aligned}
$$

where we've made the substitution $v = \frac{u}{F_0}$. If $F_L(y') \leq 0$, then we replace with zero to obtain

$$
P(y_0 \leq Y_i \leq y')\left(\mathbb{E}[Y_i|y_0 \leq Y_i \leq y'] - y_0\right) \geq -\frac{F_0^2}{f_0}\int_0 1\ln(v)\,du = \frac{F_0^2}{f_0}
$$

When $y' < y$, write the quantity of interest as:

$$
P(y' \leq Y_i \leq y_0)\left(y_0 - \mathbb{E}[Y_i|y' \leq Y_i \leq y_0]\right) = \int_{y'}^{y_0} (y_0 - y)f(y)dy = \int_{F(y')}^{F_0} (y_0 - Q(u))du
$$

This integral is decreasing in $F(y')$, so an upper bound is:

$$
\begin{aligned}
P(y' \leq Y_i \leq y_0)\left(y_0 - \mathbb{E}[Y_i|y' \leq Y_i \leq y_0]\right) &\leq \int_{F_L(y')}^{F_0} (y_0 - Q_L(u))du = -\frac{F_0}{f_0}\int_{F_L(y')}^{F_0} \ln\left(\frac{u}{F_0}\right)du \\
&= -\frac{F_0^2}{f_0}\int_{F_L(y')/F_0}^1 \ln(v)\,du \\
&= \frac{F_0 F_L(y')}{f_0}\left(\ln\left(\frac{F_L(y')}{F_0}\right) - 1\right) + \frac{F_0^2}{f_0}
\end{aligned}
$$

or simply F_0^2/f_0 when $F_L(y') \leq 0$, and a lower bound is:

$$\begin{aligned}
P(y' \leq Y_i \leq y_0)\,(y_0 - \mathbb{E}[Y_i|y' \leq Y_i \leq y_0]) &\geq \int_{F_U(y')}^{F_0} (y_0 - Q_U(u))du \\
&= \frac{1-F_0}{f_0} \int_{F_U(y')}^{F_0} \ln\left(\frac{1-u}{1-F_0}\right) du \\
&= -\frac{(1-F_0)^2}{f_0} \int_{\frac{1-F_U(y')}{1-F_0}}^{1} \ln(v)\, dv \\
&= \frac{(1-F_0)(1-F_U(y'))}{f_0} \left(\ln\left(\frac{1-F_U(y')}{1-F_0}\right) - 1\right) + \frac{(1-F_0)^2}{f_0}
\end{aligned}$$

or simply $(1-F_0)^2/f_0$ when $F_U(y') \geq 1$. □

In estimation, I censor intermediate CDF bound estimates based on he above lemma at zero and one. These constraints are not typically binding so I ignore the effect of this on asymptotic normality of the final estimators, when constructing confidence intervals.

H.10 Details of calculating wage correction terms

For the ex-post effect of the kink

Suppose that straight-time wages w^* are set according to Equation (1) for all workers, where h^* are their anticipated hours. The straight-wages that would exist absent the FLSA w_0^*, yield the same total earnings z^*, so:

$$w_0^* h^* = w^*(h^* + (\rho_1 - 1)(h^* - k)\mathbb{1}(h^* > k))$$

where $k = 40$ and $\rho_1 = 1.5$. The percentage change is thus

$$(w_0^* - w^*)/w^* = \frac{(\rho_1 - 1)(h^* - k)\mathbb{1}(h^* > k)}{h^* + (\rho_1 - 1)(h^* - k)\mathbb{1}(h^* > k)}$$

If h_{0i} is constant elasticity in the wage with elasticity $\mathcal{E}$, then we would expect

$$\frac{h_{0it} - h_{0it}^*}{h_{0it}} = 1 - \left(1 + \frac{(\rho_1 - 1)(h^* - k)\mathbb{1}(h^* > k)}{h^* + (\rho_1 - 1)(h^* - k)\mathbb{1}(h^* > k)}\right)^{\mathcal{E}}$$

Taking $h_{0it} \approx h_{1it} \approx h^*$ and integrating along the distribution of h_{1it}, we have:

$$\mathbb{E}[h_{0it} - h_{0it}^*] \approx \mathbb{E}\left[\mathbb{1}(h_{it} > k)h_{it}\left(1 - \left(1 + \frac{(\rho_1 - 1)(h_{it} - k)}{h_{it} + (\rho_1 - 1)(h_{it} - k)}\right)^{\mathcal{E}}\right)]\right]$$

which will be negative provided that $\mathcal{E} < 0$.The total ex-post effect of the kink is:

$$\mathbb{E}[h_{it} - h_{0it}^*] = \mathbb{E}[h_{it} - h_{0it}] + \mathbb{E}[h_{0it} - h_{0it}^*]$$

For a move to double-time

The straight-wages w_2^* that would exist with double time, for workers with $h^* > k$, that yield the same total earnings z^* as the actual straight wages w^* satisfy:

$$w_2^*(k + (\bar{\rho}_1 - 1)(h^* - k)) = w^*(k + (\rho_1 - 1)(h^* - k))$$

where $\bar{\rho}_1 = 2$. The percentage change is thus

$$(w_2^* - w^*)/w^* = \frac{k + (\rho_1 - 1)(h^* - k)}{k + (\bar{\rho}_1 - 1)(h^* - k)} - 1$$

Let $\bar{h}_{0i}$ be hours under a straight-time wage of w_2^*. By a similar calculation thus:

$$\mathbb{E}[\bar{h}_i^{[\bar{\rho}_1,k]} - h_{it}^{[\bar{\rho}_1,k]}] \approx \mathbb{E}\left[\mathbb{1}(h_{it} > k)h_{it}\left(\left(\frac{k + (\rho_1 - 1)(h^* - k)}{k + (\bar{\rho}_1 - 1)(h^* - k)}\right)^{\mathcal{E}} - 1\right)]\right]$$

The total effect of a move to double-time is:

$$\mathbb{E}[\bar{h}_{it}^{[\bar{\rho}_1,k]} - h_{it}] = \mathbb{E}[\bar{h}_{it}^{[\bar{\rho}_1,k]} - h_{it}^{[\bar{\rho}_1,k]}] + \mathbb{E}[h_{it}^{[\bar{\rho}_1,k]} - h_{it}]$$

The above definitions are depicted visually in Figure 21 below.

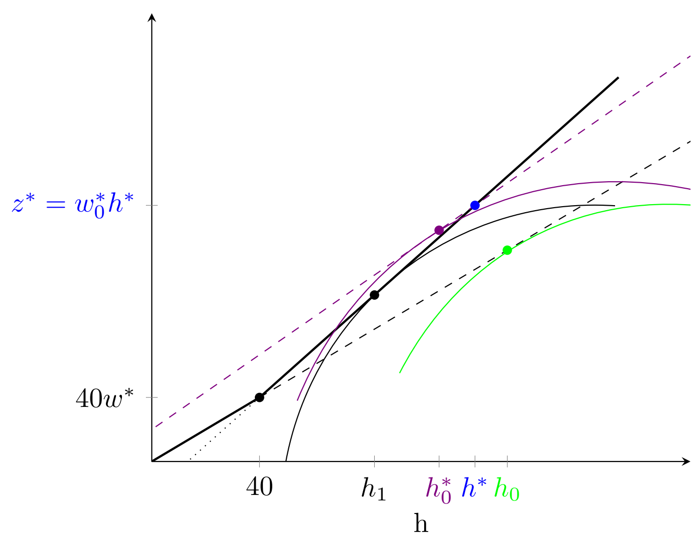


Figure 21: Depiction of h^*, h_0, h_0^* and h_1 for a single fixed worker that works overtime at h_1 hours this week. Their realized wage w^* has been set to yield earnings z^* based on anticipated hours h^* given the FLSA kink. In a world without the FLSA, the worker's wage would instead be $w_0^* = z^*/h^*$, and this week the firm would have chosen h_0^* hours, where the worker's marginal productivity this week is w_0^* (in the benchmark model). *Note:* while (z^*, h^*) is chosen jointly with employment and on the basis of anticipated productivity, choice of h_0^* is instead constrained by the contracted purple pay schedule (with the worker already hired) and on the basis of updated productivity. h_1 may differ from h^* for this same reason. In the numerical calculation h^* is approximated by h_1 – which corresponds to productivity variation being small and h^* being a credible choice given the FLSA. If credibility (the firm not wanting to renege too far on hours after hiring) were a constraint on the choice of (z^*, h^*) in the no-FLSA counterfactual, then h^* would be smaller without the FLSA, but I consider this "second-order" and do not attempt a correction here.

Changing the location of the kink

Let $\mathcal{B}_w^{[k]}$ denote bunching with the kink at location k and (a distribution of) wages denoted by w. Then the effect of moving k on bunching is

$$\mathcal{B}_{w'}^{[k']} - \mathcal{B}_w^{[k^*]} = \left(\mathcal{B}_w^{[k']} - \mathcal{B}_w^{[k^*]}\right) + \left(\mathcal{B}_{w'}^{[k']} - \mathcal{B}_w^{[k']}\right)$$

where w' are the wages that would occur with bunching at the new kink point k'. The first term has been estimated by the methods described above, with the second term representing a correction due to wage adjustment. Taking $Y_{0i} \approx Y_{1i} \approx h^*$, the straight-time wages w^* set according to Equation (1) that would change are those between k' and k^*. Consider the case $k' < k^*$. We expect wages to fall, as the overtime policy becomes more stringent, and $\left(\mathcal{B}_{w'}^{[k']} - \mathcal{B}_w^{[k']}\right)$ is only nonzero to the extent that the increase in Y_0 and Y_1 changes the mass of each in the range $[k', k^*]$. With the range $[k', k^*]$ to the left of the mode of Y_{0i}, it is most plausible that this mass will decrease. Similarly, for Y_{1i}, it is most likely that this mass will decrease, making the overall sign of $\left(\mathcal{B}_{w'}^{[k']} - \mathcal{B}_w^{[k']}\right)$ ambiguous However, since most of the adjustment should occur for workers who are typically found between k and k', we would not expect either term to be very different from zero.

Now consider the effect of average hours:

$$\mathbb{E}[Y_{w'}^{[k']} - Y_w^{[k^*]}] = \mathbb{E}[Y_w^{[k']} - Y_w^{[k^*]}] + \mathbb{E}[Y_{w'}^{[k']} - Y_w^{[k']}]$$

For a reduction in k, we would expect wages w' to be lower with $k = k'$ and hence the second term positive. This will attenuate the effects that are bounded by the methods above, holding the wages fixed at their realized levels.

Consider first the case of $k' < k^*$. Let w' be wages under the new kink point k', and assuming they adjust to keep total earnings z^* constant, wages w' will change if w^* is between k and k' as:

$$w'(k' + 0.5(h^* - k')) = w^* h^*$$

And the percentage change for these workers is thus

$$(w' - w^*)/w^* = \frac{h^*}{k' + 0.5(h^* - k')} - 1$$

$$\mathbb{E}[Y_{w'}^{[k']} - Y_w^{[k']}] \approx \mathbb{E}\left[\mathbb{1}(k' < Y_i < k^*) Y_i \left(\left(\frac{Y_i}{k' + 0.5(Y_i - k')}\right)^{\mathcal{E}} - 1\right)]\right]$$

In the case of $k' > k^*$, we will have wages change as:

$$w' h^* = w^*(k^* + 0.5(h^* - k^*))$$

w^* is between k and k'. And the percentage change for these workers is thus

$$(w' - w^*)/w^* = \frac{k^* + 0.5(h^* - k^*)}{h^*} - 1$$

$$\mathbb{E}[Y_{w'}^{[k']} - Y_{w}^{[k']}] \approx \mathbb{E}\left[\mathbb{1}(k^* < Y_i < k')Y_i\left(\left(\frac{k^* + 0.5(Y_i - k^*)}{Y_i}\right)^{\mathcal{E}} - 1\right)\right]$$

References

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.

BLOMQUIST, S., KUMAR, A., LIANG, C.-Y. and NEWEY, W. (2021). "On Bunching and Identification of the Taxable Income Elasticity". *Journal of Political Economy* 129 (8).

BLOMQUIST, S., KUMAR, A., LIANG, C.-Y. and NEWEY, W. K. (2015). "Individual heterogeneity, nonlinear budget sets and taxable income". *The Institute for Fiscal Studies Working Paper* CWP21/15.

BLOMQUIST, S. and NEWEY, W. (2017). "The Bunching Estimator Cannot Identify the Taxable Income Elasticity". *The Institute for Fiscal Studies Working Paper* CWP40/17.

BORUSYAK, K., JARAVEL, X. and SPIESS, J. (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". *arXiv:2108.12419 [econ]*.

BURDETT, K. and MORTENSEN, D. T. (1998). "Wage Differentials, Employer Size, and Unemployment". *International Economic Review* 39 (2), p. 257.

CAHUC, P. and ZYLBERBERG, A. (2004). *Labor economics.* OCLC: 265445233. Cambridge, Mass.: MIT Press.

CESARINI, D., LINDQVIST, E., NOTOWIDIGDO, M. J. and ÖSTLING, R. (2017). "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries". *American Economic Review* 107 (12), pp. 3917–46.

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.

DIAMOND, R. and PERSSON, P. (2016). *The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests.* w22207. Cambridge, MA: National Bureau of Economic Research, w22207.

DÜMBGEN, L., KOLESNYK, P. and WILKE, R. A. (2017). "Bi-log-concave distribution functions". *Journal of Statistical Planning and Inference* 184, pp. 1–17.

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.

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.

GARDNER, J. (2021). "Two-stage difierences in difierences". *Working Paper*, p. 34.

HAMERMESH, D. S. (1996). *Labor demand*. Princeton, NJ: Princeton Univ. Press.

HWANG, H., MORTENSEN, D. T. and REED, W. R. (1998). "Hedonic Wages and Labor Market Search". *Journal of Labor Economics* 16 (4), pp. 815–847.

IMBENS, G. W., RUBIN, D. B. and SACERDOTE, B. I. (2001). "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players". *The American Economic Review* 91 (4), p. 25.

KASY, M. (2022). "Who wins, who loses? Identification of the welfare impact of changing wages". *Journal of Econometrics* 226 (1), pp. 1–26.

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.

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

LEWBEL, A. and LINTON, O. (2002). "Nonparametric Censored and Truncated Regression". *Econometrica* 70 (2), pp. 765–779.

LEWIS, H. G. (1969). "Employer Interest in Employee Hours of Work". *Unpublished paper*.

MANNING, A. (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.

MILGROM, P. and ROBERTS, J. (1996). "The LeChatelier Principle". *American Economic Review* 1 (86), pp. 173–179.

SAEZ, E. (2010). "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3), pp. 180–212.

SATO, R. (1975). "The Most General Class of CES Functions". *Econometrica* 43 (5), p. 999.

TREJO, B. S. J. (1991). "The Effects of Overtime Pay Regulation on Worker Compensation". *American Economic Review* 81 (4), pp. 719–740.